स्वाध्याय स्वमन्थन स्वावलम्बन

UTTAR PRADESH RAJARSHI TANDON OPEN UNIVERSITY

(Established vide U.P. Govt. Act No. 10, of 1999)

Indira Gandhi National Open University

UP Rajarshi Tandon Open University

UGMM-02
Linear Algebra

FIRST BLOCK
Vector Spaces

Shantipuram (Sector-F), Phaphamau, Allahabad - 211013

UTTAR PRADESH
RAJARSHI TANDON OPEN UNIVERSITY

UGMM-02
LINEAR ALGEBRA

Block

1

VECTOR SPACES

LINEAR ALGEBRA

In the 19th century a lot of innovation took place in algebra. Many new systems were being studied. These included fields and vector spaces, around which linear algebra has been built up. Arthur Cayley, an English mathematician (1821-1895), was responsible for a lot of creativity in this area. Now vector spaces have become very important. One reason is that, if you want to study models of real life situations, you may need to know something about vector spaces. This and some allied concepts are also helpful in studying various aspects of physics, chemistry, economics or psychology.

For quite some time there has been a feeling among mathematical educationists all over the world that some linear algebra should be studied by any undergraduate who wants to study mathematic. Keeping this in mind we have designed the present course, assuming that you have studied mathematics in school for 12 years.

The course is divided into four blocks, as given in the programme guide. In the first block we start with recalling basic concepts like sets and functions. Then we introduce you to fields and vector spaces, the building blocks of linear algebra. In the second block we discuss linear transformations and matrices, and the close relationship between them. You will study eigenvectors and eigenvalues in Block 3. To obtain them you must familiarise yourself with determinants, which are defined and discussed in Block 3 also. The last block discusses inner product spaces and some applications of linear algebra to geometry.

After studying this course you will be in a better position to appreciate our course on abstract algebra. This course will also help you in studying our later courses on numerical analysis and differential equations.

While going through the units you will find that they have been divided into sections. Since the material in the different units is heavily interlinked, we will be doing a lot of cross-referencing. For this we will be using the notation Sec x.y to mean Section y of Unit x.

In each block we will give a block introduction, followed by a list of symbols used in the block, and then the units of the block. Every unit has exercises interspersed with the text. These exercises are meant to help you check your progress. After every exercise we leave a space for you to write your solutions in. The solutions or answers to the exercises in a unit are given at the end of the unit. After you finish going through a unit, please go back to the objectives of the unit (given at the beginning) and check if they have been achieved.

During your study of this course we will send you three assignments. They are also meant to be learning aids. We have also made a video cassette "Linear Transformations and Matrices" which you can see at your study centre. You will find the media note accompanying this programme at the end of Block 2 of this course.

We will also be making an audio programme related to this course, about the utility of linear algebra.

If you feel like reading more than what we give in our course, you may consult the following books:

1. An Introduction to Linear Algebra by V. Krishnamurthy et al., (Affiliated East-West Press).
2. University Algebra by N.S. Gopalakrishnan, (Wiley-Eastern Ltd.).

These books are available at your study centre.

NOTATIONS AND SYMBOLS

$\{x \mid x \in P\}$ $\{x : x \in P\}$	The set of all x such that x satisfies property P.
$\in$	belongs to
$\notin$	does not belong to
$\subseteq (\subsetneq)$	contained in (is properly contained in)
$\not\subseteq$	is not contained in
$A \cup B$	The union of the sets A and B
$A \cap B$	The intersection of the sets A and B
$A \setminus B$	A complement B
N	The set of natural numbers
Z	The set of integers
Q	The field of rational numbers
R	The field of real numbers
C	The field of complex numbers
R^n (C^n)	n-dimensional real space (complex space)
$\Rightarrow$	implies
$\Leftrightarrow$	implies and is implied by
iff	if and only if
$a \mid b$	a divides b
$< (\leq)$	is less than (less than or equal to)
$> (\geq)$	is greater than (greater than or equal to)
$\exists$	there exists
$\forall$	for all
$\mathbf{a} \cdot \mathbf{b}$	The scalar product of **a** and **b**
$\sum_{i=1}^{n} a_i$	$a_1 + a_2 + \ldots\ldots + a_n$
[S]	The linear span of S
$\dim_F V$	The dimension of V over **F**
$\therefore$	therefore
i.e.	that is

GREEK ALPHABETS

α	Alpha
β	Beta
γ	Gamma
δ	Delta
ε	Epsilon
θ	Theta
λ	Lambda
μ	Mu
ξ	Xi
π, Π	Pi (capital pi)
ρ	Rho
σ, Σ	Sigma (capital sigma)
τ	Tau
ϕ	Phi
ψ	Psi
ω	Omega

BLOCK 1 VECTOR SPACES

In this block we will present the basis of some mathematics that was developed in the 18th and 19th century.

We start this block with Unit 1, in which we recall the definitions of sets and related concepts. We also briefly discuss fields, an algebraic structure that you will use throughout this course.

In Unit 2 we give a detailed discussion of the two- and three-dimensional spaces R^2 and R^3. This unit shows the link between geometry and linear algebra. It leads us very naturally to a general vector space, an algebraic structure that we introduce to you in Unit 3.

In Unit 4 we dig a little deeper into the properties of the elements of a vector space. We discuss the concepts of linear independence, basis and dimension.

Since the material covered in Units 3 and 4 form the foundation of the course, we advise you to be absolutely sure of your grasp of the concepts given in them before you go any further.

UNIT 1 SETS, FUNCTIONS AND FIELDS

Structure

1.1 INTRODUCTION

This unit seeks to introduce you to the pre-requisites of linear algebra. We recall the concepts of sets, relations and functions here. These are fundamental to the study of any branch of mathematics. In particular, we study binary operations on a set, since this concept is necessary for the study of algebra. We conclude with defining a field, which is a very important algebraic structure, and give some examples of it.

Objectives

After studying this unit, you should be able to

- identify and work with sets, relations, functions and binary operations;
- recognise a field;
- give examples of finite and infinite fields.

1.2 SETS

We shall recall that the term set is used to describe any well defined collection of objects, that is, every set should be so described that given any object it should be clear whether the given object belongs to the set or not.

For instance,

a) the collection N of all natural numbers, and

b) the collection of all positive integers which divide 48 (namely, the integers 1,2,3,4,6,8,12,16,24 and 48) are well defined, and hence, are sets.

But the collection of all rich people is not a set, because there is no way of deciding whether a human being is rich or not.

If S is a set, an object a in the collection S is called an element of S. This fact is expressed in symbols as $a \in S$ (read "a is in S" or "a belongs to S"). If a is not in S, we write $a \notin S$. For example, $3 \in \mathbf{R}$, the set of real numbers. But $\sqrt{-1} \notin \mathbf{R}$.

The Greek letter epsilon, $\in$, denotes 'belongs to'

There are usually two ways of describing a set (1) Roster Method, and (2) Set Builder Method.

Roster Method: In this method, we list all the elements of the set within braces. For instance, as we have mentioned above, the collection of all positive divisors of 48 contains 1,2,3,4,6,8,12,16,24 and 48 as its elements. So this set may be written as

{1,2,3,4,6,8,12,16,24,48}.

In this description of a set, the following two conventions are followed:

Convention 1: The order in which the elements of the set are listed is not important

Convention 2: No element is written more than once: that is, every element must be written exactly once.

For example, consider the set S of all integers between $1\frac{1}{2}$ and $4\frac{1}{4}$. Obviously, these integers are 2,3 and 4. So we may write

$S = \{2,3,4\}$.

We may also write $S = \{3,2,4\}$, but we must not write $S = \{2,3,2,4\}$. Why? Isn't this what Convention 2 says?

The roster method is sometimes used to list the elements of a large set also. In this case we may not want to list all the elements of the set. We list some and give an indication of the rest of the elements. For example, the set of integers lying between 0 and 100 is $\{0,1,2,\ldots,100\}$.

Another method that we can use for describing a set is the

Set Builder Method: In this method we first try to find a property which characterises the elements of the set, that is, a property P which all elements of the set possess, and which no other objects possess. Then we describe the set as

$\{x \mid x \text{ has property } P\}$, or as

$\{x : x \text{ has property } P\}$.

This is to be read as "the set of all x such that x has property P".

For example, the set S of all integers lying between $1\frac{1}{2}$ and $4\frac{1}{4}$ can also be written as $S = \{x : x \text{ is an integer and } 1\frac{1}{2} < x < 1\frac{1}{4}\}$.

Example 1: Write the set N by the set builder method and the roster method.

Solution: By the set builder method we have the set

$N = \{x \mid x \text{ is a natural number}\}$.

By the roster method we have $N = \{1,2,3,\ldots\}$.

E E1) Write the following sets by the roster method.

$A = \{x \mid x \text{ is an integer and } 10 < x < 15\}$

$B = \{x \mid x \text{ is an even integer and } 10 < x < 15\}$

$C = \{x \mid x \text{ is a positive divisor of } 20\}$

$D = \{p/q \mid p,q \text{ integers and } 1 \le p < q \le 3\}$

E E2) Write the following sets by the set builder method:

$P = \{7,8,9\}$; $Q = \{1,2,3,5,7,11\}$; $R = \{3,6,9,\ldots\}$.

Let us now look at some sets that can be obtained from given sets.

1.2.1 Subsets, Unions, Intersections

Consider the sets $A = \{1,3,4\}$ and $B = \{1,4\}$. Here every element of B is also an element of A. In such a case, that is, when every element of B is an element of A, we say that B is a subset of A, and we write $B \subseteq A$.

It is obvious that if A is any set, then every element of A is certainly an element of A. So, for every set A, $A \subseteq A$.

Consider the sets $Q = \{1,3,5,15\}$ and $S = \{2,3,5,7\}$. Is $Q \subseteq S$? No, because not every element of Q is in S; for example, $1 \in Q$ but $1 \notin S$. Is $S \subseteq Q$? No, because, for example, $2 \in S$ but $2 \notin Q$. Therefore, there do exist pairs of sets, A and B, such that neither of them is contained in the other. This is written as $A \not\subseteq B$ and $B \not\subseteq A$ ('$\not\subseteq$' denotes 'is not a subset of'.)

∃ denotes 'there exists'

Note that if B is not a subset of A, there must be an element of B which is not an element of A. In mathematical notation this can be written as '$\exists\, x \in B$ such that $x \notin A$'.

We can now say that two sets A and B are equal (i.e., have precisely the same elements) if and only if $A \subseteq B$ and $B \subseteq A$.

E3) Which of the following statements are true?

a) $N \subseteq Z$ b) $Z \subseteq N$ c) $\{0\} \subseteq \{1,2,3\}$ d) $\{2,4,6\} \not\subseteq \{2,4,8\}$

We now give one way of obtaining a new set from two or more given sets.

Union: If we have two sets A and B, we can collect the elements of both to get a new set. This set is called their union. Formally, we define the **union of A and B** to be the set of all those elements which are in A or in B or in both A and B. The union of A and B is denoted by **A ∪ B**. Thus,

$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$.

Example 2: Find $A \cup B$ when
a) $A = \{1,2\}$ and $B = \{4,6,7\}$.
b) $A = \{1,2,3,4\}$ and $B = \{2,4,6,8\}$.

Solution: a) $A \cup B = \{1,2,4,6,7\}$.
b) $A \cup B = \{1,2,3,4,6,8\}$. Observe that 2 and 4 are in both A and B, but when we write $A \cup B$, we write these elements only once, in accordance with Convention 2 given earlier.

Can you see that, for any set A, $A \cup A = A$?

E4) Show that, if $A \subseteq C$ and $B \subseteq C$, then $A \cup B \subseteq C$.

The definition of union can be immediately extended to define the union of more than two sets. If $A_1, A_2, \ldots, A_k$ are k sets, their union $A_1 \cup A_2 \cup \ldots \cup A_k$ is the set of elements which belong to at least one of these sets. That is,

$A_1 \cup A_2 \cup \ldots \cup A_k = \{x \mid x \in A_i \text{ for some } i, i = 1,2,\ldots,k\}$

The expression $A_1 \cup A_2 \cup \ldots \cup A_k$ is often abbreviated to $\bigcup_{i=1}^{k} A_i$.

Now let us look at another way of obtaining a new set from two or more given sets.

A set having no elements is called an empty set or null set, and is denoted by ϕ, the Greek letter phi.

Intersection: If A and B are two sets, then the **Intersection of A and B** (denoted by $A \cap B$) is the set of elements common to A and B. So,

$A \cap B = \{x \mid x \in A \text{ and } x \in B\}$.

Thus, if $P = \{1,2,3,4\}$ and $Q = \{2,4,6,8\}$, then $P \cap Q = \{2,4\}$.

Can you see that, for any set A, $A \cap A = A$?

Now suppose $A = \{1,2\}$ and $B = \{4,6,7\}$. Then what is $A \cap B$? We observe that, in this case, A and B have no common elements, and so $A \cap B = \phi$, the empty set.

When the intersection of two sets is ϕ, we say that the two sets are **disjoint** (or **mutually disjoint**). For example, the two sets $\{1,4\}$ and $\{0,5,7,14\}$ are disjoint.

The definition of intersection can be extended to any number of sets. Thus, the intersection of k sets $A_1, A_2, \ldots, A_k$ is

$A_1 \cap A_2 \cap \ldots \cap A_k = \{x \mid x \in A_i \text{ for each } i = 1,2,\ldots,k\}$.

The expression $A_1 \cap A_2 \cap \ldots \cap A_k$ is abbreviated to $\bigcap_{i=1}^{k} A_i$.

Example 3: If $A \subseteq B$, what is $A \cap B$?

Solution: Since $A \subseteq B$, we know that every element of A is an element of B.

Then $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

$= \{x \mid x \in A\} = A$.

Example 4: For every set A, show that $\phi \subseteq A$.

Solution: We have already made the remark that if B is not a subset of A, there must be an element of B which is not an element of A. So if ϕ is not a subset of A, we should be able to produce an element in ϕ which is not in A. Can we do so? Obviously not! Because ϕ has no elements at all. We are therefore forced to the conclusion that $\phi \subseteq A$.

E E5) For every set A, show that $\phi \cup A = A$ and $\phi \cap A = \phi$.

E E6) State whether the following are true or false.

a) If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

b) If $A \nsubseteq B$ and $B \nsubseteq A$, then A and B are disjoint.

c) $A \nsubseteq (A \cup B)$

d) $B \subseteq (A \cup B)$

e) If $A \cup B = \phi$, then $A = B = \phi$.

E E7) Suppose $A = \{a,b,c\}$, $B = \{a,b,p,q\}$ and $C = \{a,p,r,s\}$. Find the following sets:

a) $A \cup B$, b) $B \cap C$, c) $(A \cup B) \cap C$, d) $(A \cap C) \cup (B \cap C)$.

What do you guess from your answers to (c) and (d)?

Is $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$? Check your guess by making your own choice for A, B and C

Apart from the operations of unions and intersections, there is another operation on sets, namely, the operation of taking complements.

Complements: When we are working with elements and subsets of a single set X, we say that the set X is the **universal set.** Suppose X is the universal set and $A \subseteq X$, then the set of all elements of X which are not in A is called the **complement of A** and is denoted by

A', A^c or $X \setminus A$. Thus,

$A^c = \{x \mid x \in X, x \notin A\}$.

If $X = \{a,b,p,q,r\}$ and $A = \{a,p,q\}$, then clearly $A^c = \{b,r\}$.

E8) Why are the following statements true?

a) A and A^c are disjoint, i.e., $A \cap A^c = \phi$.

b) $A \cup A^c = X$, where X is the universal set.

c) $(A^c)^c = A$.

Certain properties of the complements of sets have been stated as **De Morgan's Laws.** We give them as a theorem.

Theorem 1: If A and B are subsets of X, then

a) $(A \cup B)^c = A^c \cap B^c$

b) $(A \cap B)^c = A^c \cup B^c$

(In words, 'complement of union is intersection of complements' and 'complement of intersection is union of complements'.)

Proof: a) Two sets P and Q are equal, if and only if $P \subseteq Q$ and $Q \subseteq P$, that is, if and only if $x \in P \Rightarrow x \in Q$ and $x \in Q \Rightarrow x \in P$.

'$\Rightarrow$' denotes 'implies'

Thus, to prove (a), we must prove that

$x \in (A \cup B)^c \Rightarrow x \in A^c \cap B^c$ and $x \in A^c \cap B^c \Rightarrow x \in (A \cup B)^c$.

Now

$x \in (A \cup B)^c \Rightarrow x \notin A \cup B$

$\Rightarrow x \notin A$ and $x \notin B$

$\Rightarrow x \in A^c$ and $x \in B^c$

$\Rightarrow x \in A^c \cap B^c$

Conversely,

$x \in A^c \cap B^c \Rightarrow x \in A^c$ and $x \in B^c$

$\Rightarrow x \notin A$ and $x \notin B$

$\Rightarrow x \notin A \cup B$

$\Rightarrow x \in (A \cup B)^c$

Note that in both parts of the proof, the various steps are the same but only in reverse order. When this is the case, both parts can be combined as follows.

'$\Leftrightarrow$' denotes 'implies and is implied by' or 'if and only if'.

$x \in (A \cup B)^c \Leftrightarrow x \notin A \cup B$

$\Leftrightarrow x \notin A$ and $x \notin B$

$\Leftrightarrow x \in A^c$ and $x \in B^c$

$\Leftrightarrow x \in A^c \cap B^c$

E E9) Try and prove (b) (of Theorem 1) now.

So far we have looked at sets algebraically. Now let us look at them pictorially.

1.2.2 Venn Diagrams

Some results about sets can be easily understood and visualised by using Venn diagrams, named after the English logician John Venn (1834-1923). In a Venn diagram the universal set is usually represented by a rectangle and its subsets by circles or other closed figures in its interior. For example, if A, B and C are subsets of X, this fact is represented by the following diagram (Fig. 1).

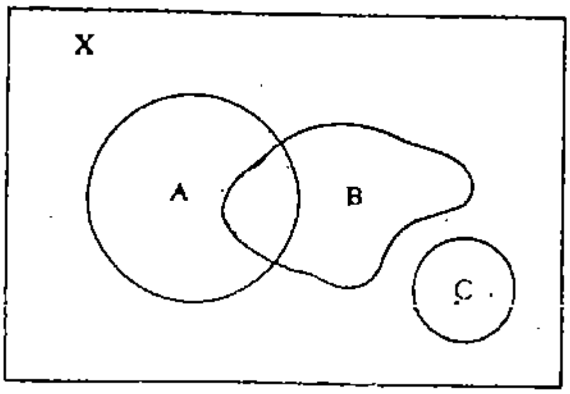


Fig. 1

The idea is that points in the interior of the rectangle represent the elements of X and the points in the interior of the closed figures, A, B and C represent the elements of A, B and C, respectively. Notice that the subsets of X can be of any shape.

If $X = \{a,b,c,p,q,r,s\}$, $A = \{a,b,p,r\}$ and $B = \{p,q,r\}$, then this can be represented by the following Venn diagram (Fig. 2).

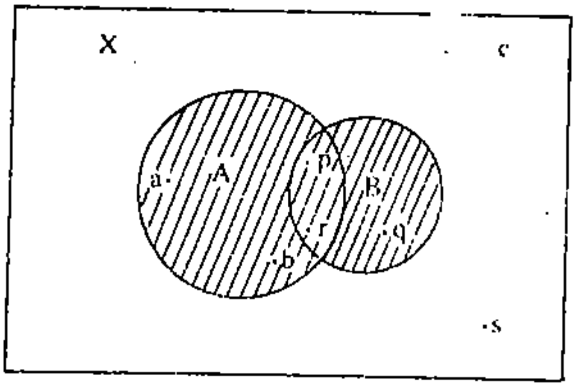


Fig. 2

Then, A U B is the shaded portion in Fig. 2, and $(A \cup B)^c$ is the unshaded portion of the diagram.

E10) Use Venn diagrams to demonstrate the truth of the following results. Here A,B,C are subsets of X.

a) $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

b) $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

We will now talk of the product of sets, of which the coordinate system is a special case.

1.3 CARTESIAN PRODUCT OF SETS

An interesting set that can be formed from two given sets is their **Cartesian product**, named after the French philosopher and mathematician Rene Descartes (1596 – 1650) He also invented the Cartesian coordinate system.

Rene Descartes

Let A and B be two sets. Consider the pair (a,b), in which the first element is from A and the second from B. Then (a,b) is called an **ordered pair**. In an ordered pair, the order in which the two elements are written is important. Thus, (a,b) and (b,a) are **different ordered pairs.** Two ordered pairs (a,b) and (c,d) are said to be equal or same, if $a = c$ and $b = d$.

Definition: The Cartesian product $A \times B$, of the sets A and B, is the set of all possible ordered pairs (a,b), where $a \in A$, $b \in B$.

That is, $A \times B = \{(a,b) : a \in A, b \in B\}$.

For example, if $A = \{1,2,3\}$, $B = \{4,6\}$, then

$A \times B = \{(1,4), (1,6), (2,4), (2,6), (3,4), (3,6)\}$

Also note that

$B \times A = \{(4,1), (4,2), (4,3), (6,1), (6,2), (6,3)\}$, and $A\times B \neq B\times A$.

We can also define the Cartesian product of more than two sets in a similar way. Thus, if **R** is the set of all real numbers, then

$R \times R = \{(a_1, a_2) : a_1 \in R, a_2 \in R\}$,

$R \times R \times R = \{(a_1, a_2, a_3) : a_i \in R\}$,

and so on. It is customary to write R^2 for R × R and R^n for R × .... × R (n times).

Since every point in a plane has two coordinates, x and y, and every ordered pair (x,y) of real numbers defines the coordinates of a point in the plane, we say $\mathbf{R}^2$ represents a plane. Thus, $\mathbf{R}^2$ is the Cartesian product of the x-axis and the y-axis. In the same way $\mathbf{R}^3$ represents three-dimensional space, and $\mathbf{R}^n$ represents n-dimensional space, for any $n \geq 1$. Note that **R** represents a line

E E11) If A = {2,5}, B = {2,3}, find A×B, B×A, A×A.

E E12) If A × B = {(7,2), (7,3), (7,4), (2,2), (2,3), (2,4)}, determine A and B.

E E13) Prove that (A ∪ B) × C = (A×C) ∪ (B×C).

Let us now look at subsets of certain Cartesian products.

1.4 RELATIONS

You are already familiar with the concept of a relationship between people. For example, a parent-child relationship exists between A and B if and only if A is a parent of B or B is a parent of A.

In mathematics, a relation R on a set S is a relationship between the elements of S. If $a \in S$ is related to $b \in S$ by means of this relation, we write a R b, or $(a,b) \in R$. From the latter notation we see that $R \subseteq S\times S$. And this is exactly how a (binary) relation on a set is defined.

Definition: A relation R on a set S is a subset of $S \times S$.

For example, if N is the set of natural numbers and R is the relation 'is a multiple of', then 15R5, but not 5R15. That is, $(15,5) \in R$ but $(5,15) \notin R$. Here $R \subseteq N \times N$.

Again, if Q is the set of all rational numbers and R is the relation 'is greater than', then 3R2 (because $3 > 2$). In fact, for any number $n > 1$, $nR(n-1)$.

E14) Let N be the set of all natural numbers and R the relation 'is a divisor of' on the set N. State whether the following are true or false.

a) 2 R 3

b) $n R n, \forall n \in N$

c) $n R m$ and $m R n \Rightarrow m = n$

We now look at some particular kinds of relations.

Definition: A relation R defined on a set S is said to be

i) **reflexive** if we have $a R a \ \forall a \in S$.

ii) **symmetric** if $a R b \Rightarrow b R a \ \forall a, b \in S$.

iii) **transitive** if $a R b$ and $b R c \Rightarrow a R c \ \forall a, b, c \in S$.

To get used to these concepts, consider the following examples.

Example 5: Let N be the set of all natural numbers. We define the relation R on N as follows:

$a R b$ if and only if $a > b$.

Determine whether R is reflexive, symmetric and transitive.

Solution: Since $a > a$ is not true, so $a R a$ is not true. Hence, R is not reflexive. If $a > b$ then certainly $b > a$ is not true. That is, $a R b$ does not imply $b R a$. Hence, R is not symmetric.

Since $a > b$ and $b > c$ implies $a > c$, we find that $a R b$, $b R c$ implies $a R c$. Thus, R is transitive.

E15) The relation $R \subseteq N \times N$ is defined by $(a,b) \in R$ iff 5 divides $(a-b)$. Is R reflexive, symmetric or transitive?

The relationship in E 15 is reflexive, symmetric and transitive. Such a relation is called an **equivalence relation.**

A very important property of an equivalence relation on a set S is that it divides S into a number of mutually disjoint subsets, that is, it **partitions** S. Let us see how this happens.

Let R be an equivalence relation on the set S. Let $a \in S$. Then the set $S_a = \{b \mid b \in S, aRb\}$ is called the **equivalence class** of a in S. It is just the set of elements in S which are related to a. For instance, for R given in E15, what is the equivalence class of 1?

This is

$$N_1 = \{n \mid 1Rn, n \in N\}$$
$$= \{n \mid n \in N \text{ and } 5 \text{ divides } 1-n\}$$
$$= \{n \mid n \in N \text{ and } 5 \text{ divides } n-1\}$$
$$= \{1, 6, 11, 16, 21, \ldots\}$$

Similarly,

$N_2 = \{n \mid n \in N \text{ and } 5 \text{ divides } n-2\}$

$= \{2,7,12,17,22,\ldots\}$

$N_3 = \{3,8,13,18,23,\ldots\}$

$N_4 = \{4,9,14,19,24,\ldots\}$

$N_5 = \{5,10,15,20,25,\ldots\}$

$N_6 = \{1,6,11,16,21,\ldots\}$

$N_7 = \{2,7,12,17,22,\ldots\}$, etc.

Note that

i) $N = N_1 \cup N_2 \cup N_3 \cup N_4 \cup N_5$, and the sets on the right hand side are mutually disjoint.

ii) N_1 and N_6 are not disjoint. In fact, $N_1 = N_6$. Similarly $N_2 = N_7$, and so on.

These observations will be proved in general in the following theorem.

Theorem 2: Let R be an equivalence relation on a set S. For $a \in S$, let S_a denote the equivalence class of a. Then,

a) $S = \bigcup_{a \in S} S_a$

b) If $a, b \in S$ then $S_a \cap S_b = \phi$ or $S_a = S_b$

Proof: a) Since $S_a \subseteq S \; \forall \; a \in S$, $\bigcup_{a \in S} S_a \subseteq S$ (see E 4).

Conversely, let $x \in S$. Then, $x \in S_x$ (as x R x is true.) And S_x is one of the sets in the collection $\{S_a \mid a \in S\}$, whose union is $\bigcup_{a \in S} S_a$.

Hence, $x \in \bigcup_{a \in S} S_a$. So $S \subseteq \bigcup_{a \in S} S_a$

Thus, $S \subseteq \bigcup_{a \in S} S_a$ and $\bigcup_{a \in S} S_a \subseteq S$, proving (a).

b) Suppose $S_a \cap S_b \neq \phi$. Let $x \in S_a \cap S_b$.

Then, $x \in S_a$ and $x \in S_b$.

$\Rightarrow$ aRx and bRx

$\Rightarrow$ aRx and xRb (since R is symmetric)

$\Rightarrow$ aRb (since R is transitive)

Using this we shall prove that $S_a = S_b$. For this, take $y \in S_a$. Then aRy, which is the same as yRa. We have also shown that aRb. This gives us yRb, since R is transitive That is, bRy, which means that $y \in S_b$. Thus, $S_a \subseteq S_b$. Similarly, it can be proved that $S_b \subseteq S_a$. Thus, $S_a = S_b$, and (b) is proved.

Note that, in the above theorem, because of (b), distinct sets on the right hand side of (a) are all disjoint from one another. Therefore, (a) expresses S as a union of mutually disjoint subsets of S; that is we have a partition of S into equivalence classes.

E E16) Show that 'aRb if and only if a = b' is an equivalence relation on **Z**. What is Z_1?

E E17) Let $S = A \cup B \cup C$, where A,B,C are mutually disjoint, and R is the relation defined on S by:
aRb if whenever $a \in A, b \in A$ or whenever $a \in B$, $b \in B$, or whenever $a \in C$, $b \in C$.

Prove that R is an equivalence relation on S. For $b \in B$, what is S_b?

In the next section we discuss a familiar concept that also leads to some relations.

1.5 FUNCTIONS

Recall that a function f from a set A to a set B is a rule which associates with every element of A exactly one element of B. This is written as $f : A \to B$. If f associates with $a \in A$, the element b of B, we write $f(a) = b$. A is called the **domain** of f and the set $\{f(a) \mid a \in A\}$ is called the **range** of f. The range of f is a subset of B. B is called the **co-domain** of f.

Note that
i) For each element of A, we associate some element of B.
ii) For each element of A, we associate only one element of B.
iii) Two or more elements of A could be associated with the same element of B.

For example, let $A = \{1,2,3\}$, $B = \{1,2,3,4,5,6,7,8,9,10\}$. Define $f : A \to B$ by $f(1) = 1, f(2) = 4, f(3) = 9$. Then f is a function. In this case we can also write $f(x) = x^2$ for each $x \in A$. The domain of f is A and the range is $\{1,4,9\}$.

We could also have written the definition of f as $f : A \to B : f(x) = x^2$. We will often use this notation for defining any function.

If we define $g : A \to B$ by $g(1) = 1, g(2) = 1, g(3) = 4$, then g is also a function. The domain of g remains the same, namely, A. The range of g is $\{1,4\}$.

A function $f : A \to B$ is said to be one-one (or **injective**) if different elements of A are associated with different elements of B, i.e., if $a_1, a_2 \in A$ and $a_1 \neq a_2$, then $f(a_1) \neq f(a_2)$. In the foregoing examples, the function f is one-one. The function g is not one-one because 1 and 2 are distinct elements of A, but $g(1) = g(2)$.

'f is one-one' can also be written as 'f is 1-1'.

Now consider another example of sets and functions.
Let $A = \{1,2,3\}$, $B = \{p,q,r\}$. Let $f: A \to B$ be defined by $f(1) = q$, $f(2) = r$, $f(3) = p$. Then f is a function. Here the range of f = B = co-domain of f. This is an example of an onto function, as you shall see.

Definition: A function $f: A \to B$ is said to be **onto** (or **surjective**) if the range of f is B, i.e., if, for each $b \in B$, there is an $a \in A$ such that $f(a) = b$.

If a function is both one-one and onto it is called **bijective**. The example of an onto function given above is also 1-1, and hence, bijective.

E18) Let $f: \mathbf{N} \to \mathbf{N}$ be defined by $f(n) = n+5$. Prove that f is one-one but not onto.

E19) Let $f: \mathbf{Z} \to \mathbf{Z}$ be defined by $f(n) = n+5$. Prove that f is both one-one and onto.

E20) Let $A = \{1, -1, 2, 3\}$. If $f: A \to \mathbf{R}$ is defined by $f(x) = x^2 - 5x + 6$, find the range of f.

Two functions that you will often come across are

i) the identity function $I_A: A \to A: I_A(a) = a \ \forall \ a \in A$,

ii) the constant function $f: A \to B: f(a) = c \ \forall \ a \in A$, where c is a fixed element of B.

E21) a) Can you show that the identity function is bijective?

b) What must X be like for the constant function $f: X \to \{c\}$ to be injective?

Now let us see what we mean by the composition of two or more functions.

1.5.1 Composition of Functions

If $f: A \to B$ and $g: C \to D$ are functions and if the range of f is a subset of C, there is a natural way of combining g and f to yield a new function $h: A \to D$. For each $x \in A$, $h(x)$ is defined by the formula $h(x) = g(f(x))$. (Note that $f(x)$ is in the range of f, so that $f(x) \in C$. Therefore, $g(f(x))$ is defined and is an element of D.) This function h is called the **composition of g and f** and is written as **gof**. The domain of gof is A and its codomain is **D**.

Example 6: Let $f: \mathbf{R} \to \mathbf{R}$ and $g: \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = x^2$ and $g(x) = x+1$. What is gof? What is fog?

Solution: We observe that the range of f is a subset of R, the domain of g. Therefore, g∘f is defined. By definition, $\forall x \in R$,

$g \circ f(x) = g(f(x)) = f(x) + 1 = x^2 + 1$.
Now, let us find f∘g. Again, it is easy to see that f∘g is defined.
$\forall x \in R$, $f \circ g(x) = f(g(x)) = (g(x))^2 = (x+1)^2$.
Thus, $g \circ f \neq f \circ g$.

f : A → B and g : C → D are equal if A = C and f(a) = g(a) ∀ a ∈ A.

Example 7: Let A = {1,2,3}, B = {p,q,r} and C = {x,y}. Let f: A → B be defined by f(1) = p, f(2) = p, f(3) = r. Let g: B → C be defined by g(p) = x, g(q) = y, g(r) = y. Determine if f∘g and g∘f can be defined.

Solution: For f∘g to be defined, it is necessary that the range of g should be a subset of the domain of f. In this case, the range of g is C and the domain of f is A. As C is not a subset of A, f∘g cannot be defined.

Since the range of f, which is {p,r}, is a subset of B, the domain of g, we see that g∘f is defined.

Also g∘f : A → C is such that

$g \circ f(1) = g(f(1)) = g(p) = x$
$g \circ f(2) = g(f(2)) = g(p) = x$
$g \circ f(3) = g(f(3)) = g(r) = y$

In this example note that g is surjective, and so is g∘f.

E22) In each of the following questions, both f and g are functions from R to R. Define f∘g and g∘f, wherever meaningful.

a) $f(x) = 5x$, $g(x) = x + 5$
b) $f(x) = 5x$, $g(x) = x/5$
c) $f(x) = x^3$, $g(x) = \sin x + 3$
d) $f(x) = |x|$, $g(x) = x^2$.

Remark: Functions can lead to relations. How can this happen? Given a function f: A → B, can you define a relation? What about $R \subseteq A \times B$, where $(a,b) \in R$ iff $b = f(a)$? This is a relation that arises from f.

We now come to a theorem which shows us that the identity function behaves like the number $1 \in R$ does for multiplication. That is, if we take the composition of any function f with the suitable identity function, we get the same function f.

Theorem 3: For every function $f:A \to A$, we have $f \circ I_A = f$ and $I_A \circ f = f$.

Proof: Since both f and I_A are defined from A to A, both the compositions $f \circ I_A$ and $I_A \circ f$ are defined. Moreover, $\forall x \in A$,

$f \circ I_A(x) = f(I_A(x)) = f(x)$, so $f \circ I_A = f$.

Also, $\forall x \in A$,

$I_A \circ f(x) = I_A(f(x)) = f(x)$, so $I_A \circ f = f$.

On the lines of this theorem you can try the next exercise.

E E23) If B is any set and $g:B \to A$, prove that $I_A \circ g = g$ and $g \circ I_B = g$.

In the case of real numbers, you know that given any real number $x \neq 0$, $\exists y \neq 0$ such that $xy = 1$. y is called the inverse of x. Similarly, we define an inverse function for a given function.

Definition: Let $f: A \to B$ be a given function. If there exists a function $g: B \to A$ such that $f \circ g = I_B$ and $g \circ f = I_A$, then we say that g is the inverse of f, and we write $g = f^{-1}$.

For example, consider $f: R \to R$ defined by $f(x) = x+3$. If we define $g:R \to R$ by $g(x) = x-3$, then $f \circ g(x) = f(g(x)) = g(x)+3 = (x-3)+3 = x \; \forall x \in R$. Hence, $f \circ g = I_R$. Similarly, $g \circ f = I_R$ (verify). So $g=f^{-1}$

Note that, in the above example, f adds 3 to x and g does the opposite—it subtracts 3 form x. Thus, the key to finding the inverse of a given function is : try to retrieve x from (x).

For example let $f:R \to R$ be defined by $f(x) = 3x + 5$. How can we retrieve x from $3x + 5$? The answer is "first subtract 5 and then divide by 3". So we try $g(x) = \frac{x-5}{3}$. And we find

$$g \circ f(x) = g(f(x)) = \frac{f(x)-5}{3} = \frac{(3x+5)-5}{3} = x$$

Also, $f \circ g(x) = 3(g(x)) + 5 = 3\frac{(x-5)}{3} + 5 = x \; \forall x \in R$.

E E24) What is the inverse of $f:R \to R$: $f(x) = \frac{x}{3}$?

Do all functions have an inverse? No, as the following example shows.

Example 8: Let $f: R \to R$ be given by $f(x) = 1 \; \forall x \in R$. What is the inverse of f?

Solution: If f has an inverse $g: R \to R$ we have $f \circ g = I_R$, i.e., $\forall x \in R$, $f \circ g(x) = x$. Now take $x = 5$. We should have $f \circ g(5) = 5$, i.e., $f(g(5)) = 5$. But $f(g(5)) = 1$, since $f(x) = 1 \; \forall x$. We reach a contradiction. Therefore, f has no inverse.

In view of this example, we naturally ask for necessary and sufficient conditions for f to have an inverse. The answer is given by the following theorem.

Theorem 4: A function $f:A \to B$ has an inverse if and only if f is bijective.

Proof: First suppose f is bijective. We shall define a function $g:B \to A$ and then prove that $g = f^{-1}$.

Let $b \in B$. Since f is onto, there is some $a \in A$ such that $f(a) = b$, and, as f is one-one, there is only one such $a \in A$. We take this unique element a of A as g(b). That is, given $b \in B$, we define $g(b) = a$, where $f(a) = b$.

Note that, since f is onto, $B = \{f(a)|a \in A\}$. Then, we are simply defining $g:B \to A$ by $g(f(a)) = a$. This automatically ensures that $g \circ f = I_A$.

Now, for this g, we prove that $g = f^{-1}$. Let $a \in A$. Then $g \circ f(a) = g(f(a)) = a$, by the definition of g, so that $g \circ f = I_A$.

Next, let $b \in B$. Then, if $g(b) = a$, we must have
$f(a) = b$ (by the definition of g), so $f \circ g(b) = f(g(b)) = f(a) = b$.

Hence, $f \circ g = I_B$.
This proves that $g = f^{-1}$.

Conversely, suppose f has an inverse and let $g = f^{-1}$ We must prove that f is one-one and onto.
Suppose $f(a_1) = f(a_2)$ then $g(f(a_1)) = g(f(a_2))$.
$\Rightarrow g \circ f(a_1) = g \circ f(a_2)$
$\Rightarrow a_1 = a_2$, because $g \circ f = I_A$.
So f is one-one.

Finally, given $b \in B$, we have $f \circ g = I_B$, so that $f \circ g(b) = I_B(b) = b$, i.e., $f(g(b)) = b$. So, given $b \in B$, there is $g(b) \in A$ such that $f(g(b)) = b$. That is, f is onto.

Hence the theorem is proved.

E E25) Consider the following functions from **R** to **R**. For each determine whether it has an inverse and, when the inverse exists, find it.
a) $f(x) = x^2$
b) $f(x) = 0$
c) $f(x) = 11x + 7$

We now come to a particular kind of function, namely, a binary operation.

1.5.2 Binary Operation

You are familiar with the operations of addition and multiplication on the set of real numbers. Addition is a function which associates with $(a,b) \in \mathbf{R}^2$ the element $a + b$ of **R**. So, it is a function from $\mathbf{R} \times \mathbf{R}$ to **R**. Can you see that multiplication is also a function from $\mathbf{R} \times \mathbf{R}$ to **R**? These functions can be performed on any two elements of **R**. They are examples of binary operations, which we now define.

Definition: A binary operation on a non-empty set S is a function from $S \times S$ to S.

Thus, a binary operation on S associates a unique element in S to each pair of elements in S. The word 'binary' means involving pairs. It is customary to denote a binary operation by a symbol such as +, . , 0, *, etc.

As mentioned earlier, + and × are binary operations on **R**.
Another example is $*: \mathbf{R} \times \mathbf{R} \to \mathbf{R}: a * b = \frac{a+b}{2}$

Some binary operations can have special properties which we now define.

Definition : A binary operation $*$ on a set S is said to be
a) closed on a subset T of S if $t_1 * t_2 \in T \ \forall \ t_1, t_2 \in T$.
b) commutative if $a * b = b * a \ \forall \ a, b \in S$.
c) associative if $(a*b)*c = a*(b*c) \ \forall \ a,b,c \in S$.

For example, the operations of addition and multiplication on **R** are commutative as well as associative. But, subtraction is neither commutative nor associative on **R**. Why? Is $a-b = b-a$, or $(a-b)-c = a-(b-c) \ \forall \ a,b,c \in \mathbf{R}$? No. For example, $1-2 \neq 2-1$ and $(1-2)-3 \neq 1-(2-3)$. Also subtraction is not closed on $\mathbf{N} \subseteq \mathbf{R}$, because $1 \in \mathbf{N}, 2 \in \mathbf{N}$ but $1-2 \notin \mathbf{N}$.

Note that a binary operation on S is always closed on S, but may not be closed on a subset of S.

In calculations you must have often used the fact that $a(b+c) = ab + ac$ and $(b+c)a = ba + ca \ \forall \ a,b,c \in \mathbf{R}$. We say that multiplication distributes over addition in **R**. In general, we have the following definition.

Definition: If o and $*$ are two binary operations on a set S, we say that $*$ **is distributive over o** if $\forall \ a,b,c \in S$, we have
$a*(b \circ c) = (a*b)\circ(a*c)$, and $(b \circ c)*a = (b*a)\circ(c*a)$.

Example 9: Let $a*b = \frac{a+b}{2} \ \forall \ a,b \in \mathbf{R}$. Prove that the operation of multiplication in **R** distributes over $*$.

Solution: We have to see whether $a(b*c) = ab * ac$ and $(b*c)\,a = ba * ca$.

$$\text{Now } a(b*c) = a\frac{(b+c)}{2} = \frac{ab+ac}{2} = ab * ac.$$

$$\text{Also } (b*c)a = \frac{(b+c)}{2}a = \frac{ba+ca}{2} = ba * ca.$$

Hence, multiplication is distributive over $*$.

Now, go back to E10. What does it say? It says that the intersection of sets distributes over the union of sets and the union of sets distributes over the intersection of sets.

Let us now look deeper at some binary operations. You know that, for any $a \in \mathbf{R}$, $a + 0 = a$ and $0+a = a$ and $a+(-a) = (-a)+a = 0$. We say that 0 is the identity element for addition and $(-a)$ is the negative, or additive inverse, of a. In general, we have the following definition.

Definition: Let $*$ be a binary operation on a set S. If there is an element $e \in S$ such that $\forall \ a \in S$, $a * e = a$ and $e * a = a$, then e is called **an identity element for** $*$.

For $a \in S$, we say that $b \in S$ is an **inverse** of a, if $a * b = e$ and $b * a = e$. In this case, we usually write $b = a^{-1}$

In the following theorem we will prove the uniqueness of the identity element for $*$, and the uniqueness of the inverse of an element with respect to $*$, if it exists.

Theorem 5: Let $*$ be a binary operation on a set S. Then
a) if $*$ has an identity element, it must be unique.
b) if $*$ is associative and $s \in S$ has an inverse with respect to $*$, it must be unique.

Proof: a) Suppose e and e' are both identity elements for $*$.
Then $e = e * e'$, since e' is an identity element
$\quad = e'$, since e is an identity element.
That is, $e = e'$. Hence, the identity element is unique.

b) Suppose there exist $a, b \in S$ such that $s * a = e = a * s$ and $s * b = e = b * s$, e being the identity element for $*$. Then

$$a = a*e = a*(s*b)$$
$$= (a*s)*b, \text{ since } * \text{ is associative}$$
$$= e*b = b$$

That is, $a = b$.
Hence, the inverse of s is unique.

This theorem allows us to use the identity element and the inverse, henceforth.

Example 10: If the binary operation $\oplus: \mathbf{R} \times \mathbf{R} \to \mathbf{R}$ is defined by $a \oplus b = a+b-1$, prove that $\oplus$ has an identity. If $x \in \mathbf{R}$, determine the inverse of x with respect to $\oplus$, if it exists.

Solution: We are looking for some $e \in \mathbf{R}$ such that $a \oplus e = a = e \oplus a \ \forall \ a \in \mathbf{R}$. Now, $a \oplus e = a + e - 1$. So we want $e \in \mathbf{R}$ such that $a + e - 1 = a \ \forall \ a \in \mathbf{R}$.
Obviously, $e = 1$ will satisfy this. Also, $1 \oplus a = a \ \forall \ a \in \mathbf{R}$. Hence, 1 is the identity element of $\oplus$.

For $x \in \mathbf{R}$, if b is the inverse of x, we should have $b \oplus x = 1$,
i.e., $b + x - 1 = 1$, so $b = 2 - x$.
Indeed, $(2-x) \oplus x = (2-x) + x - 1 = 1$, and $x \oplus (2-x) = x + 2 - x - 1 = 1$.
So $x^{-1} = 2 - x$.

E26) For the following binary operations defined on R, determine whether they are commutative, associative or have identity elements.
a) $x \oplus y = x + y - 5$.
b) $x * y = 2(x + y)$
c) $x \Delta y = \frac{x-y}{2}$

Now that you are familiar with sets and binary operations we will study sets with particular types of operations. This course is built on such sets.

1.6 FIELDS

You must be familiar with the sets
Q, of all rational numbers
R, of all real numbers
C, of all complex numbers.

In this section you will discover that these sets are examples of fields. Q and R were known to Euclid, way back in 300 B.C. The complex number field, C, is relatively new. It was developed in the 18th century.

All fields need not be infinite sets. You will also come across fields with only a finite number of elements. These finite fields were studied by Gauss in his book Disquisitiones Arithmeticae.

In the following definition we will talk of properties of the binary operations denoted by '+' and '·'. Do not confuse these with the usual addition and multiplication in **R** (though these operations in R do satisfy the properties given, as you can check for yourself as we go along).

Definition: Let + and · be two binary operations on a non-empty set F. The set F is called a field if the following 9 properties hold $\forall$ a,b,c $\in$ F.

A1) + is associative: (a+b)+c = a+(b+c)

A2) $\exists$ an identity element with respect to +, denoted by 0:
a+0 = 0+a = (0 is called the zero element).

A set F for which A1–A3 hold is called a group with respect to +

A3) Every element of F has an inverse with respect to +: for any a $\in$ F, $\exists$ b $\in$ F such that a+b = 0 = b+a. b is written as (−a) and is called the inverse of a with respect to +.

A4) + is commutative: a + b = b + a.

M1) · is associative: (a·b)·c = a·(b·c)

M2) $\exists$ an identity element with respect to·, denoted by e:
a·e = e·a = a.

F is called a ring with respect to + and · if it satisfies A1–A4, M1 and D.

M3) Every element of F \ {0} has an inverse with respect to · : for any a $\in$ F\{0} $\exists$ b $\in$ F \ {0} such that a.b = e = b·a. (b is written as a^{-1})

M4) · is commutative: a·b = b·a

D) · distributes over + :
(a+b)·c = a·c + b·c and a·(b+c) = a·b + a·c

The operation that satisfies A1 – A4 is called **addition**, and its inverse operation is called subtraction. The other binary operation is called **multiplication**, and its inverse operation is called division. Thus, subtraction and division are defined by
$a - b = a+(-b)$, and $a \div b = a \cdot b^{-1}$ for $b \neq 0$.
Note that a field is closed under the basic operations of addition, subtraction and multiplication. The set of non-zero elements of a field are closed under division. Can you see that both Q and R are fields? You just need to check that they satisfy the 9 properties for the usual operations of addition and multiplication. The system **Z** is not a field because, for example, 2 $\in$ **Z** does not have a multiplicative inverse in **Z**. This violates property M3.

E27) Show that the system C of complex numbers is also a field, the operations being given by
(a+ib) + (c+id) = (a+c) + i(b+d), and
(a+ib) · (c+id) = ac − bd + i (bc+ad) $\forall$ a,b,c,d $\in$ **R**. ($i = \sqrt{-1}$.)

An important property of every field is expressed in the following result.

Theorem 6: If F is a field, then $\forall$ a $\in$ F, a·0 = 0.

Proof: Let a·0 = b.

Then b = a·0 = a(0 + 0) (because 0 + 0 = 0)
= a·0 + a·0 (distributive property)
= b + b

That is, b = b + b

Hence 0 = b + (−b) = (b+b) + (−b)
= b+(b+(−b)) (Associative property)
= b + 0
= b

Thus, b = 0, i.e., a·0 = 0

So far we have only given examples of infinite fields. Now we give an example of a finite field.

Example 11: On the set Z_3 = {0,1,2} we define the binary operations $\oplus$ and $\odot$ as follows:

x $\oplus$ y = remainder left on dividing x + y by 3.
x $\odot$ y = remainder left on dividing x · y by 3.
$\forall$ x, y $\in$ Z_3.

Show that Z_3 is a field. It is called the **field of integers modulo 3.**

Solution: It can be easily verified that both the operations are commutative and associative. 0 and 1 are the additive and multiplicative identities respectively. The additive inverse of 0,1,2, are 0,2 and 1, respectively. The multiplicative inverses of 1 and 2 are 1 and 2, respectively. You can also verify that multiplication is distributive over addition. So Z_3 is a field. Note that Z_3 is a finite field since it only has 3 elements.

In general, given any prime number p, we get a finite field Z_p. The underlying set of Z_p is {0,1,2,..,p−1}. The binary operations on Z_p are $\oplus$ and $\odot$ defined as follows:

x $\oplus$ y = remainder left on dividing x + y by p.

x $\odot$ y = remainder left on dividing x·y by p.

$\forall$ x,y $\in$ Z_p.

These fields are called **prime fields.**

E28) If R = {a / b | a,b $\in$ Z, b odd}, is R a field?

Before ending this section we will define an important trait of a field, namely, its characteristic.

Definition: If, for a field F, $\exists$n $\in$ N such that na = 0 $\forall$ a $\in$ F, then the least such positive integer n is called the **characteristic** of the Field F.

If no such positive integer exists we say that the field is of characteristic 0.

Example 12: What are the characteristics of

a) Q, R, C?

b) the prime field Z_p, for any prime number p?

Solution: a) For any $n \in N$ and $x \in Q$, $nx = 0 \Rightarrow x = 0$. $\therefore$ the characteristic of Q is zero. Similarly, the characteristics of R and C are zero.

b) The characteristic of any prime field Z_p is p. Why? Well, what happens if you take an element $x \in Z_p$, and divide px by p? The remainder is zero. That is, $p \odot x = 0$, for any $x \in Z_p$. Also, if you take any natural number m, $0 < m < p$, then $m \odot 1 = m \neq 0$. Therefore, p is the least positive integer such that $p \odot x = 0 \ \forall \ x \in Z_p$. This tells us that the characteristic of Z_p is p.

It can be proved that if a field is not of characteristic zero then its characteristic has to be a prime number.

When you go to the next unit you will realise how important it is to be thoroughly familiar with fields. Do make sure that you are quite at ease with this section. Now let us briefly go through the points brought up in this unit.

1.7 SUMMARY

We conclude by summarising what we have covered in this unit. We have

1) studied the concepts of sets, subsets, complements, unions and intersections of sets.
2) shown you how to represent sets by Venn diagrams.
3) defined the Cartesian product of sets, as well as relations and equivalence relations on a set.
4) defined the notions of a function, composition of functions and inverse functions.
5) studied the possible properties of binary operations on a set.
6) defined and seen many examples of fields, both infinite and finite.
7) defined the characteristic of a field.

1.8 SOLUTIONS/ANSWERS

E1) $A = \{11, 12, 13, 14\}$ $\qquad C = \{1, 2, 4, 5, 10, 20\}$

$B = \{12, 14\}$ $\qquad D = \{1/2, 1/3, 2/3\}$

E2) $P = \{x \mid x \text{ is an integer and } 6 < x < 10\}$

$= \{x \mid x \text{ is an integer and } 7 \leq x \leq 9\}$

$Q = \{x \mid x = 1 \text{ or } x \text{ is a prime number less than } 12\}$

(A prime number is a number whose only factors are 1 or itself.)

$R = \{x \mid x \text{ is a multiple of } 3\}$

E3) (a) and (d)

E4) For $x \in A \cup B$, we have $x \in A$ or $x \in B$. In either case $x \in C$. Therefore, $A \cup B \subseteq C$.

E5) $\phi \cup A = \{x \mid x \in \phi \text{ or } x \in A\}$

$= \{x \mid x \in A\}$, since ϕ has no elements.

$= A$

$\phi \cap A = \phi$, since $\phi \subseteq A$.

E6) a) True b) False c) False d) True e) True

E7) a) $\{a, b, c, p, q\}$ $\qquad$ b) $\{a, p\}$ $\qquad$ c) $\{a, p\}$ $\qquad$ d) $\{a, p\}$

$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$ always, as you will see in E10.

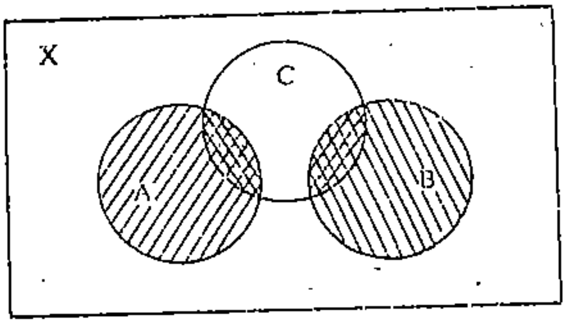
E8) a) Since $x \in A$ if and only if $x \notin A^c$, we find that A and A^c are disjoint.

b) For any $x \in X$, $x \in A$ or $x \in A^c$. Therefore, $A \cup A^c = X$.

c) Let $x \in (A^c)^c$. Then $x \notin A^c$, so that $x \in A$. Thus,
$(A^c)^c \subseteq A$. On the other hand, if $x \in A$, then $x \notin A^c$.
Hence, $x \in (A^c)^c$. $\therefore A \subseteq (A^c)^c$. $\therefore$, by the definition of equality of sets we find
$(A^c)^c = A$.

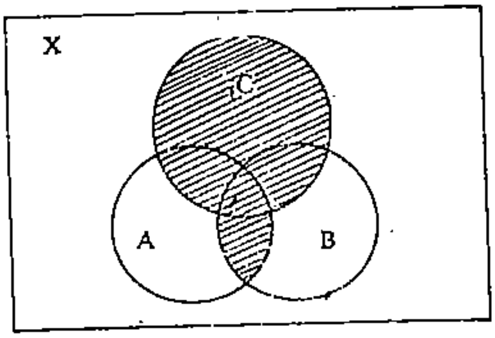
E9) $x \in (A \cap B)^c \iff x \notin A \cap B \iff x \notin A$ or $x \notin B \iff x \in A^c$ or $x \in B^c$
$\iff x \in A^c \cup B^c$.

E10) a)



$A \cup B$ is all the shaded portion. $(A \cup B) \cap C$ is the double shaded portion. As you can see from the figure, this is the same as the union of $A \cap C$ (the double shaded portion in A) and $B \cap C$ (the double shaded portion in B).

b)



$(A \cap B) \cup C$ is the shaded portion. From the figure you can see that it is the same as the set $(A \cup C) \cap (B \cup C)$.

E11) $A \times B = \{(2,2), (5,2), (2,3), (5,3)\}$
$B \times A = \{(2,2), (2,5), (3,2), (3,5)\}$
$A \times A = \{(2,2), (2,5), (5,2), (5,5)\}$

E12) Since the first element in each pair has to be in A, we get
$A = \{7,2\}$. Similarly, $B = \{2,3,4\}$.

E13) $(x,y) \in (A \cup B) \times C \iff x \in A \cup B$ and $y \in C$
$\iff x \in A$ or $x \in B$ and $y \in C$
$\iff (x,y) \in A \times C$ or $(x,y) \in B \times C$
$\iff (x,y) \in (A \times C) \cup (B \times C)$

E14) a) F b) T c) T

E15) R is reflexive since $5 \mid a - a = 0$, for any $a \in N$.
R is symmetric because, if $5 \mid a-b$, then $5 \mid b-a$, for any $a,b \in N$.
R is transitive because, if $5 \mid a-b$ and $5 \mid b-c$, then
$5 \mid (a-b) + (b-c)$, that is, $5 \mid a-c$, for any $a,b,c \in N$.

E16) R is reflexive, since $a\,R\,a \;\forall\, a \in Z$.
R is symmetric, since $a\,R\,b \Rightarrow b\,R\,a \;\forall\, a,b \in Z$.
R is transitive, since $a\,R\,b,\ b\,R\,c \Rightarrow a\,R\,c \;\forall\, a,b,c \in Z$
$Z_1 = \{x \in Z \mid x\,R\,1\} = \{1\}$

E17) R is clearly reflexive and symmetric.
Now, if a R b and $a \in A$, then $b \in A$. Again, if b R c, then, since $b \in A$, we get $c \in A$. So we find that whenever $a \in A$, $c \in A$. Similarly, if $a \in B$, $c \in B$ and if $a \in C$, $c \in C$. Thus aRb, bRc $\Rightarrow$ a R c. That is, R is transitive.
For $b \in B$, $S_b = \{x \in S \mid x R b\} = \{x \in S \mid x \in B\} = B$.

E18) Let $m, n \in N$ such that $f(m) = f(n)$. Then $m+5 = n+5$. Therefore, $m = n$. This means that $m \neq n \Rightarrow f(m) \neq f(n)$. Therefore, f is 1 – 1. f is not onto because there is no $n \in N$ such that $f(n) = 1$. Why? Well, if $f(n) = 1$, then $n+5 = 1$, and hence, $n = -4 \notin N$.

E19) f is 1 – 1, just as shown in E18.
Now f is onto because given any $z \in Z$, $\exists\, z-5 \in Z$ such that $f(z-5) = z$.

E20) The range of $f = \{f(x) \mid x \in A\}$
$= \{f(1), f(-1), f(2), f(3)\}$
$= \{2, 12, 0\}$

E21) a) $I_A: A \to A$ is 1 – 1 (since $a_1 \neq a_2 \Rightarrow I_A(a_1) \neq I_A(a_2)$), and is onto (since the range of I_A is A).
b) Suppose X has at least two elements, say x and y. Then $f(x) = c = f(y)$, but $x \neq y$. This means that f is not 1 – 1, which is a contradiction. Therefore, X also has to be a singleton, that is, have only one element, if f is to be 1 – 1.

E22) a) Both fog and gof can be defined.
$f \circ g\,(x) = f(g(x)) = 5g(x) = 5(x+5)\ \forall\, x \in R$.
$g \circ f\,(x) = g(f(x)) = f(x) + 5 = 5x + 5\ \forall\, x \in R$
Note that $f \circ g \neq g \circ f$.
b) $(f \circ g)(x) = 5(x/5) = x\ \forall\, x \in R$.
$(g \circ f)(x) = 5x/5 = x\ \forall\, x \in R$.
In this case $f \circ g = g \circ f$.
c) $(f \circ g)(x) = (\sin x + 3)^3$ and $(g \circ f)(x) = \sin x^3 + 3$.
d) $(f \circ g)(x) = |x|^2$ and $(g \circ f)(x) = |x|^2$.
In this case $f \circ g = g \circ f$.

E23) Since $I_A: A \to A$, $I_A \circ g$ is defined. Similarly, $g \circ I_B$ is defined. Now, $(I_A \circ g)(b) = I_A(g(b)) = g(b)\ \forall\, b \in B$.
$\therefore$, $I_A \circ g = g$.
Similarly, $g \circ I_B = g$.

E24) Define $g: R \to R: g(x) = 3x$. Then $f \circ g = I_R = g \circ f$.

E25) a) f is not 1 – 1 since $f(-1) = f(1)$. $\therefore$, f is not bijective and f^{-1} does not exist.
b) f is not onto, and hence f^{-1} does not exist.
c) f is bijective, and hence f^{-1} exists. In fact $f^{-1}: R \to R: f^{-1}(x) = \frac{x-7}{11}$.

E26) a) Since $x \oplus y = y \oplus x$ for any $x, y \in R$, $\oplus$ is commutative.
Since $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ for any $x, y, z \in R$, $\oplus$ is associative.
Since $x \oplus 5 = 5 \oplus x = x$, we get 5 to be the identity element for $\oplus$.
b) $*$ is commutative, not associative and has no identity element.
c) Δ is neither commutative nor associative, and has no identity element.

E27) C is a field because '+' satisfies A1 – A4, the zero being $0 + i0 = 0$, and the inverse of $a + ib$ being $(-a) + i(-b)$. '·' satisfies M1 – M4, the identity being $1 + i0 = 1$ and the inverse of $(a + ib)$ being $\frac{a - ib}{a^2 + b^2}$. D is also satisfied.

E28) Over here + and · are the usual + and · in Q. Therefore, A1 – A4 are satisfied, the zero being 0/1 (or 0/b for any odd b!). M1, M2, M4 and D are also satisfied, the multiplicative identity being 1/1. But M3 is not satisfied, since 2/1 has no inverse in R.

UNIT 2 TWO- AND THREE-DIMENSIONAL SPACES

Structure

2.1 INTRODUCTION

This unit gives the basic connection between linear algebra and geometry. Linear algebra is built up around the concept of a vector. In this unit we shall assume that you know some Euclidean plane geometry, and introduce the concept of vectors in a geometric way. For this, we begin by studying vectors in two - and three- dimensional spaces. These are called plane vectors and space vectors, respectively.

Vectors were first introduced in physics as entities which have both a measure and a definite direction (such as force, velocity, etc.). The properties of vectors were later abstracted and studied in mathematics.

Here we shall introduce a vector as a directed line segment which has length as well as a direction. Since vectors are line segments, we shall be able to define angles between vectors, perpendicular (or orthogonal) vectors, and so on.

We shall then use all this knowledge to study some aspects of the geometry of space.

Since the concepts given in this unit will be generalised in future units, you must study this unit thoroughly.

Objectives

After studying this unit, you should be able to

- define a vector and calculate its magnitude and direction;
- obtain the angle between two vectors;
- perform the operations of addition and scalar multiplication on plane vectors as well as space vectors;
- obtain the scalar product of two plane (or space) vectors;
- express a vector as a linear combination of a set of vectors that form an orthonormal basis;
- solve simple problems involving the vector equations of a line, a plane and a sphere.

2.2 PLANE AND SPACE VECTORS

How would you find out the position of a point in a plane? You would choose a set of coordinate axes and fix the point by its x and y coordinates (see Fig. 1).

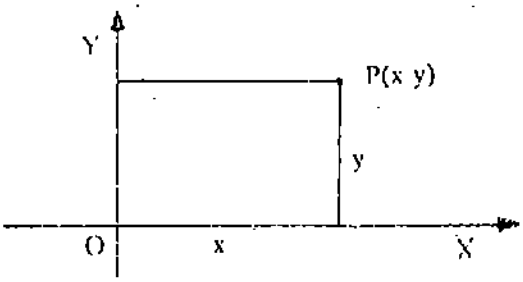
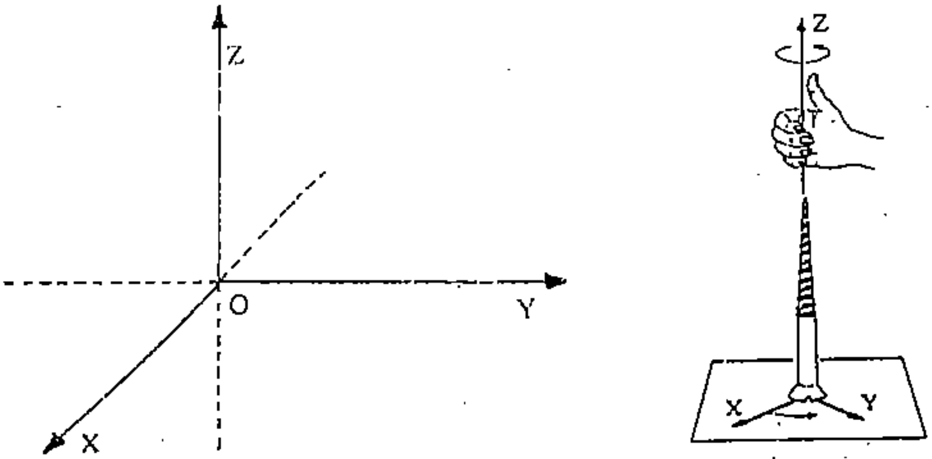


Fig. 1

Similarly, to pinpoint the position of a point in three-dimensional space, we have to give three numbers. To do this, we take three mutually perpendicular lines (axes) in space which intersect in a point O (see Fig. 2(a)). O is called the origin. The positive directions, OX, OY and OZ on those lines are so chosen that if a right-handed screw (Fig. 2(b)) placed at O is rotated from OX to OY, it moves in the direction of OZ.



(a) Fig. 2 (b)

To find the coordinates of any point P in space, we take the foot of the perpendicular from P on the plane XOY (Fig. 3). Call it M. Let the coordinates of M in the plane XOY be (x,y) and the length of MP be |z|. Then the coordinates of P are (x,y,z), where |z| is the length of MP, z is positive or negative according as MP is in the positive direction OZ or not.

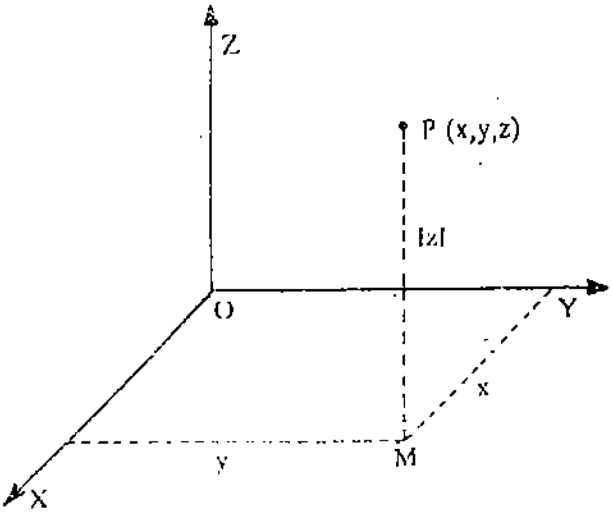


Fig. 3

So, for each point P in space, there is an ordered triple (x,y,z) of real numbers, i.e., an element of $\mathbf{R}^3$ (see Unit 1). Conversely, given an ordered triple of real numbers, we can easily find a point P in the space whose coordinates are the given triple. So there is a one-one correspondence between the space and the set $\mathbf{R}^3$. For this reason, the three-dimensional space is often denoted by the symbol $\mathbf{R}^3$. For a similar reason a plane denoted by $\mathbf{R}^2$, and a line by $\mathbf{R}$.

In R^2 or R^3 we come across entities which have magnitude and direction. They are called vectors. The word 'vector' comes from a Latin word that means 'to carry'. Let us see what the mathematical definition of a vector is.

Definition: A vector in R^2 or R^3 is a directed line segment $\overrightarrow{AB}$ with an initial point A and a terminal point B. Its **length**, or **magnitude**, is the distance between A and B, and is denoted by $|\overrightarrow{AB}|$. Every vector $\overrightarrow{AB}$ has a direction, which is from A to B. In Fig. 4, $\overrightarrow{AB}$, $\overrightarrow{CD}$, $\overrightarrow{OE}$, $\overrightarrow{CF}$ are examples of vectors with different directions.

$|\overrightarrow{AB}|$ is read as 'modulus of $\overrightarrow{AB}$'

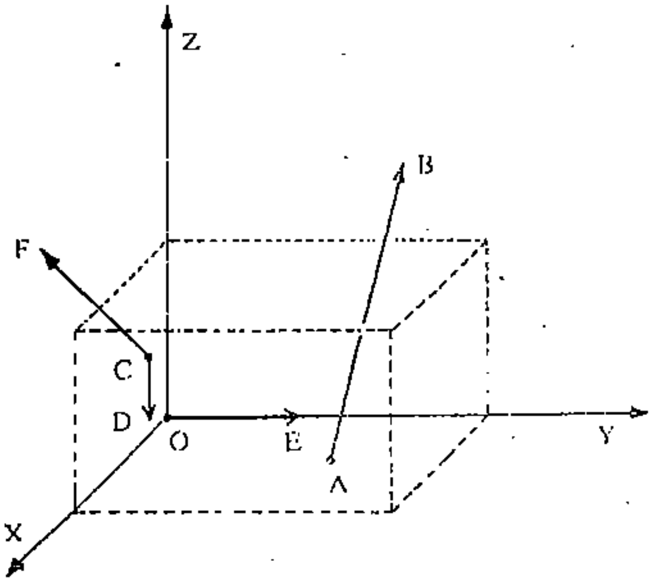


Fig. 4

Now, if $\overrightarrow{AB}$ is a plane vector and the coordinates of A are (a_1, a_2) and of B are (b_1, b_2), then $|\overrightarrow{AB}| = |\overrightarrow{BA}| = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2}$. Similarly, if $A = (a_1,a_2,a_3)$ and $B = (b_1,b_2,b_3)$ are two points in R^3, then the length of the space vector $\overrightarrow{AB}$ is $|\overrightarrow{AB}| = |\overrightarrow{BA}| = \sqrt{(a_1-b_1)^2 + (a_2-b_2)^2 + (a_3-b_3)^2}$.

The vector $\overrightarrow{AB}$ is called a **unit vector** if $|\overrightarrow{AB}| = 1$.

Definition: Two (plane or space) vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are called **parallel** if the lines AB and CD are parallel lines. If the lines AB and CD coincide, then AB and CD are said to be **in the same line.**

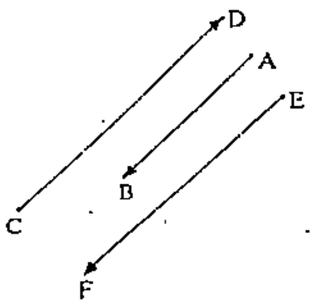


Fig. 5

From Fig. 5, you can see that two parallel vectors or two vectors in the same line may have the same direction or opposite directions. Also note that parallel vectors need not have the same length.

Definition: If $\overrightarrow{AB}$ and $\overrightarrow{CD}$ have the same length and the same direction, we say $\overrightarrow{AB}$ is **equivalent to** $\overrightarrow{CD}$. If A and C coincide, and B and D coincide, then we say $\overrightarrow{AB}$ and $\overrightarrow{CD}$ **are equal.**

Note that equivalent vectors have the same magnitude and direction but may have different initial and terminal points. In geometric applications of vectors, the initial and terminal points do not play any significant part. What is important is the magnitude and direction of a vector. Therefore, we may regard two equivalent vectors as equal vectors. This means that we are free to change the initial point of a vector (but not its magnitude or direction).

Because of this, we shall always agree to let the origin, O, be the initial point of all our vectors. That is, given any vector $\overrightarrow{AB}$, we shall represent it by the equivalent vector

$\overrightarrow{OP}$, for which $|\overrightarrow{OP}| = |\overrightarrow{AB}|$ and $\overrightarrow{OP}$ and $\overrightarrow{AB}$ have the same directions (see Fig.6). Then the terminal point, P, completely determines the vector. That is, two different points P and Q, in R^3 (or R^2) will give us two different vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$.

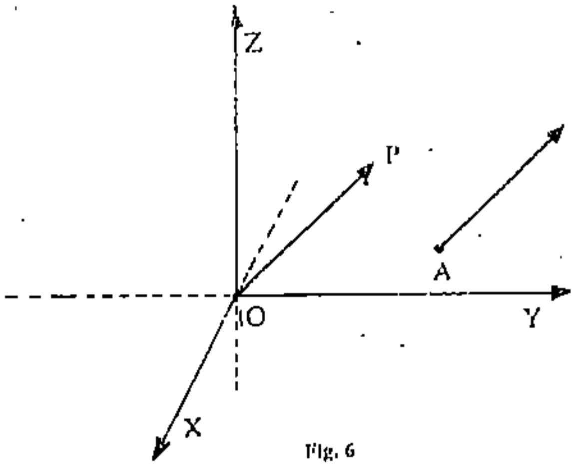


Fig. 6

E1) In Fig. 4 we have drawn 4 vectors. Draw the vectors which are equivalent to them and have O as their initial points.

As we have noted, a vector in $\mathbf{R}^2$ or $\mathbf{R}^3$ is completely determined if its terminal point is known. There is a 1 – 1 correspondence between the vectors in $\mathbf{R}^2$ (or $\mathbf{R}^3$) and the points in $\mathbf{R}^2$ (or $\mathbf{R}^3$). This correspondence allows us to make the following definition.

Definition: a) A **plane vector** is an ordered pair (a_1,a_2) of real numbers,

b) A **space vector** is an ordered triple (a_1,a_2,a_3) of real numbers.

Note that we are not making any distinction between a point P in the plane (or $P(a_1,a_2,a_3)$ in space) and the vector $\overrightarrow{OP}$ in R^2 (or R^3).

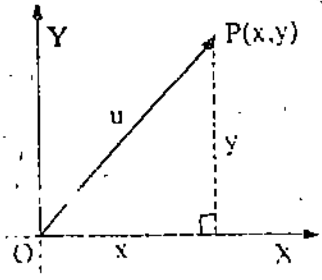


Fig. 7

We may often use a single letter **u** or **v** for a vector. Of course, **u** or **v** shall mean a pair or a triple of real numbers, depending on whether we are talking about $\mathbf{R}^2$ or $\mathbf{R}^3$. For example, $\mathbf{u} = (1,2)$; $\mathbf{v} = (0.5,-3)$, etc.

Definition: The vector (0,0) in the plane, and the vector (0,0,0) in space are called the zero vectors in R^2 and R^3, respectively.

Now, if $\mathbf{u} = (x,y)$, then can we obtain its magnitude in terms of x and y? Yes, we can. Its magnitude in given by $|\mathbf{u}| = \sqrt{x^2+y^2}$, as you can see from Fig.7 (and applying the Pythagoras Theorem!)

Similarly, if $v = (x,y,z)$, then $|v| = \sqrt{x^2+y^2+z^2}$.

Let us consider the following examples.

i) If $u = (5,12)$, $|u| = \sqrt{5^2+12^2} = 13$

ii) If $u = (-6,1)$, $|u| = \sqrt{(-6)^2+1^2} = \sqrt{37}$

iii) If $v = (1,2,-1)$, then $|v| = \sqrt{1^2+2^2+(-1)^2} = \sqrt{6}$

iv) If $v = (1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})$, then **v** is a unit vector because $|\mathbf{v}| = 1$.

v) If $w = (1/3, 1/2, \sqrt{23/36})$, then w is a unit vector because $|w| = 1$.

As we have mentioned earlier, two vectors in $\mathbf{R}^2$ are equal if their terminal points coincide (since their initial points are assumed to be at the origin). Thus, in the language of ordered pairs and triplets, we can give the following definition.

Definition: Two plane vectors (a_1,a_2) and (b_1,b_2) are said to be equal if $a_1 = b_1$ and $a_2 = b_2$. Similarly, two space vectors (a_1,a_2,a_3) and (b_1,b_2,b_3) are said to be equal if $a_1 = b_1$, $a_2 = b_2$, $a_3 = b_3$.

For example, $(a,b) = (2,3)$ if and only if $a = 2$ and $b = 3$. Also $(x,y,1) = (2,3,a)$ if and only if $x = 2$, $y = 3$ and $a = 1$.

E E2) Fill in the blanks:

a) $(2,0) = (x,y) \Rightarrow x =$ .... and $y =$ ....

b) $(1,2) = (2,1)$ is a .... statement.

c) $(1,2,3) = (1,2,z) \Rightarrow z =$ ....

Now that you have got used to plane and space vectors we go ahead and define some operations on these vectors.

2.3 OPERATIONS ON VECTORS

You are familiar with binary operations on R (Unit 1). We use these to define operations on the vectors of $\mathbf{R}^2$ and $\mathbf{R}^3$.

2.3.1 Addition

Two vectors in $\mathbf{R}^2$ can be added by considering each as an ordered pair, rather than as a directed line segment. The advantage is that we can easily extend this definition to vectors in $\mathbf{R}^3$.

Definition: The addition of two plane vectors (x_1,y_1) and (x_2,y_2) is defined by
$(x_1,y_1) + (x_2,y_2) = (x_1+x_2, y_1+y_2)$.
Similarly, the addition of two space vectors (x_1,y_1,z_1) and (x_2,y_2,z_2) is defined by
$(x_1,y_1,z_1) + (x_2,y_2,z_2) = (x_1+x_2, y_1+y_2, z_1+z_2)$.

The geometric interpretation of addition in $\mathbf{R}^2$ is easy to see. The sum of two vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$, in $\mathbf{R}^2$, is the vector $\overrightarrow{OR}$, where OR is the diagonal of the parallelogram whose adjacent sides are OP and OQ (Fig.8). Note that $\overrightarrow{QR}$ is equivalent to $\overrightarrow{OP}$.

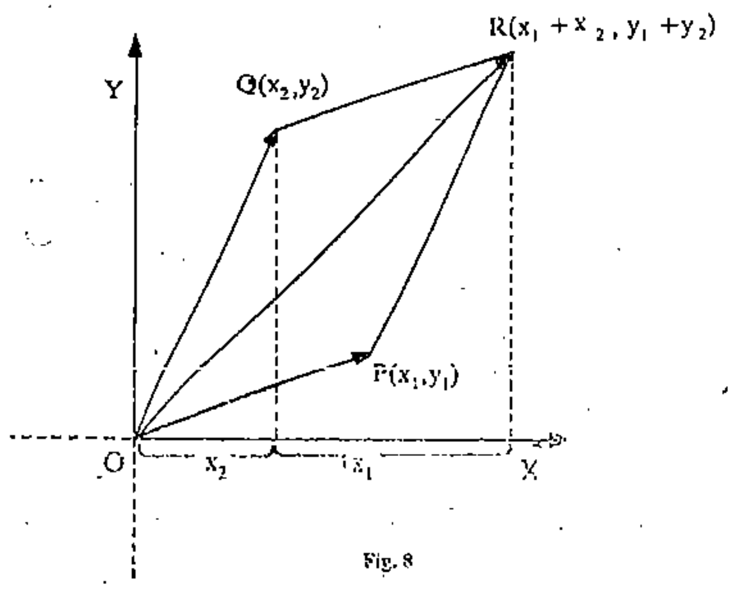


Fig. 8

Let us look at an example of addition in $\mathbf{R}^2$ and $\mathbf{R}^3$

Example 1: Find the sum of
a) $(4,-3)$ and $(0,1)$, b) $(1,-1,2)$ and $(-1,2,-5)$.

Solution: a) $(4,-3)+(0,1)=(4+0,\ -3+1)=(4,-2)$

b) $(1,-1,2)+(-1,2,-5)=(1-1,\ -1+2,\ 2-5)=(0,1,-3)$

It is obvious that the sum of two plane (or space) vectors is a plane (or space) vector, so that vector addition is a binary operation on the set of plane vectors and on the set of space vectors. The set of all space vectors, R^3, satisfies the following properties with respect to the operation of vector addition. For any (a_1,a_2,a_3), (b_1,b_2,b_3), (c_1,c_2,c_3) in R^3

i) **vector addition is associative:**

$(a_1,a_2,a_3)+\{(b_1,b_2,b_3)+(c_1,c_2,c_3)\}$
$=(a_1,a_2,a_3)+(b_1+c_1,\ b_2+c_2,\ b_3+c_3)$
$=(a_1+(b_1+c_1),\ a_2+(b_2+c_2),\ a_3+(b_3+c_3))$
$=((a_1+b_1)+c_1,\ (a_2+b_2)+c_2,\ (a_3+b_3)+c_3)$
$=\{(a_1,a_2,a_3)+(b_1,b_2,b_3)\}+(c_1,c_2,c_3)$

Note that in the above proof we have made use of the fact that the a_i's, b_i's, c_i's are real numbers, and that, for real numbers, addition is associative.

ii) **vector addition is commutative:**

$(a_1,a_2,a_3)+(b_1,b_2,b_3)$
$=(a_1+b_1,\ a_2+b_2,\ a_3+b_3)$
$=(b_1+a_1,\ b_2+a_2,\ b_3+a_3)$
$=(b_1,b_2,b_3)+(a_1,a_2,a_3)$

iii) **identity element exists for vector addition:**

Consider the vector $(0,0,0)$. We have $(a_1,a_2,a_3)+(0,0,0)=(a_1+0,\ a_2+0,\ a_3+0)=(a_1,a_2,a_3)$. Similarly, $(0,0,0)+(a_1,a_2,a_3)=(a_1,a_2,a_3)$. So $(0,0,0)$ is the identity element for vector addition. We denote this vector by **0**. (Now you know why **0** is called the zero vector!)

iv) **every space vector has an inverse with respect to addition:**

Given (a_1,a_2,a_3), consider $(-a_1,-a_2,-a_3)$. Then clearly
$(a_1,a_2,a_3)+(-a_1,-a_2,-a_3)=(0,0,0)$ and
$(-a_1,-a_2,-a_3)+(a_1,a_2,a_3)=(0,0,0)$.
So the (additive) inverse of (a_1,a_2,a_3) is $(-a_1,-a_2,-a_3)$. That is, if $\mathbf{n}=(a_1,a_2,a_3)$, then $-\mathbf{n}=(-a_1,-a_2,-a_3)$.

E E3) Show that properties (i) – (iv) hold good for $\mathbf{R}^2$.

Now that we have discussed the properties of vector addition, we are ready to define another operation on vectors.

2.3.2 Scalar Multiplication

We now consider the multiplication of any pair or triple by a real number.

'Scalar' means number.

Definition: If $\alpha \in R$ and (a_1,a_2) is a plane vector, we define the multiplication of (a_1,a_2) by the scalar α to be the plane vector $(\alpha a_1, \alpha a_2)$, i.e.,

$\alpha (a_1,a_2) = (\alpha a_1, \alpha a_2)$

Similarly, $\alpha (a_1,a_2,a_3) = (\alpha a_1, \alpha a_2, \alpha a_3)$.

What does scalar multiplication mean geometrically? To understand this we take $\overrightarrow{OP}$, a vector in R^2, and $\alpha \in R$. Then $\alpha \cdot \overrightarrow{OP}$ is a vector whose length is $|\alpha|\,|\overrightarrow{OP}|$ and whose direction is the same as that of $\overrightarrow{OP}$, if $\alpha > 0$, and opposite to that of $\overrightarrow{OP}$, if $\alpha < 0$. For example, Fig. 9 shows us 3 vectors, $\overrightarrow{OP}$, $\overrightarrow{OQ}$ and $\overrightarrow{OR}$, in R^2. Here, $\overrightarrow{OQ} = 2\,\overrightarrow{OP}$ and $\overrightarrow{OR} = -(1/2)\overrightarrow{OP}$.

$|\alpha| = \begin{cases} \alpha, \text{ if } \alpha \geq 0 \\ -\alpha, \text{ if } \alpha < 0 \end{cases}$

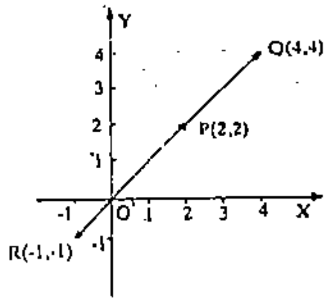


Fig. 9

Now, for any plane vector $u = (a_1,a_2)$, and for all $\alpha \in R$, we will algebraically show that $|\alpha u| = |\alpha|\,|u|$.

Since $\alpha u = (\alpha a_1, \alpha a_2)$, we get $|\alpha u| = \sqrt{\alpha^2 a_1^2 + \alpha^2 a_2^2}$

$= |\alpha| \sqrt{a_1^2 + a_2^2} = |\alpha|\,|u|$

E4) For a space vector u, prove that $|\alpha u| = |\alpha|\,|u|$, $\alpha \in R$.

Now, for any plane (or space) vector v, we define $-v$ to be $(-1)\,v$. Then $u - v = u + (-v)$, for any two plane or space vectors u and v. Thus we have defined subtraction with the help of scalar multiplication.

We now give, with proofs, 5 properties of plane vectors, related to scalar multiplication. For this, let $\alpha, \beta \in R$ and $u = (a_1,a_2)$, $v = (b_1,b_2)$ be any two plane vectors. Then

i) $\alpha (u+v) = \alpha u + \alpha v$

Scalar multiplication distributes over vector addition.

Proof: $\alpha (u+v) = \alpha [(a_1,a_2) + (b_1,b_2)]$

$= \alpha (a_1 + b_1, a_2 + b_2)$

$= (\alpha (a_1+b_1), \alpha (a_2+b_2))$

$= (\alpha a_1 + \alpha b_1, \alpha a_2 + \alpha b_2)$

$= (\alpha a_1, \alpha a_2) + (\alpha b_1, \alpha b_2)$

$= \alpha(a_1,a_2) + \alpha(b_1,b_2)$

$= \alpha u + \alpha v$

ii) $(\alpha + \beta)\, u = \alpha u + \beta u$

Proof: $(\alpha + \beta)\, u = (\alpha + \beta)\,(a_1, a_2)$

$= ((\alpha + \beta)\, a_1, (\alpha + \beta)\, a_2)$

$= (\alpha a_1 + \beta a_1, \alpha a_2 + \beta a_2)$

$= (\alpha a_1, \alpha a_2) + (\beta a_1, \beta a_2)$

$= \alpha(a_1,a_2) + \beta(a_1,a_2)$

$= \alpha u + \beta u$

iii) $\alpha(\beta u) = (\alpha\beta)u$

Proof: $\alpha(\beta u) = \alpha(\beta(a_1,a_2)) = \alpha(\beta a_1, \beta a_2)$
$= (\alpha\beta a_1, \alpha\beta a_2) = \alpha\beta(a_1,a_2)$
$= (\alpha\beta)u$

Similarly,
$\beta(\alpha u) = (\beta\alpha)u = (\alpha\beta)u$

iv) $1 \cdot u = u$

Proof: $1 \cdot u = 1(a_1,a_2) = (1a_1, 1a_2) = (a_1,a_2) = u$

v) $0 \cdot u = 0$, the zero vector in R^2.

Proof: $0 \cdot u = 0(a_1,a_2) = (0 \cdot a_1, 0 \cdot a_2) = (0,0) = 0$

E E5) Prove that the properties (i) to (v) given above also hold for the set of all space vectors.

Now that you are familiar with the operations of addition and scalar multiplication of vectors, we introduce the concept of linear combinations. (You will study more about this in Unit 3.)

Definition: A plane (or space) vector x is said to be a **linear combination** of the non-zero plane (or space) vectors $u_1, u_2, \ldots, u_n$ if there exist scalars $\alpha_1, \ldots, \alpha_n$, which are not all zero, such that $x = \alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_n u_n$.

For example, the vector (3,5) is a linear combination of the vectors (1,0) and (0,1) because (3,5) = 3(1,0) + 5(0,1). Similarly, (1,1,2) is a linear combination of the vectors (1,0,0), (0,1,0) and (0,0,1) because (1,1,2) = (1,0,0) + (0,1,0) + 2(0,0,1).

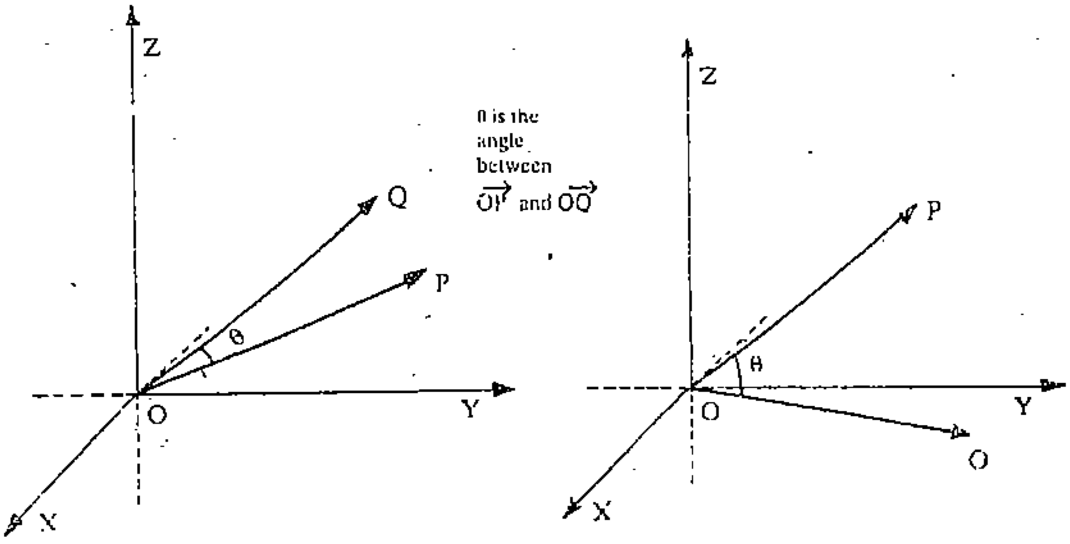
E E6) Show that every vector $(a,b) \in R^2$ is a linear combination of the vectors (1,0) and (0,1).

We end this section with mentioning that the set of all plane vectors, along with the operations of vector addition and scalar multiplication defined above, forms an algebraic structure called a vector space. (We will define the term 'vector space' in Unit 3.) Similarly, the set of all space vectors, along with vector addition and scalar multiplication defined above, forms a vector space.

Let us now look at one way of multiplying two vectors.

2.4 SCALAR PRODUCT

You know that every vector has a direction. Thus, it makes sense to speak about the angle between two vectors. You must have learnt, in Euclidean geometry, that any two intersecting lines determine a plane. Thus, given any two distinct non-zero vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$, we get a plane in which these two vectors lie. Then, the angle between $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ is the radian measure of $\angle POQ$ which is interior to $\triangle POQ$ in this plane (see Fig. 10).



(a) (b)

Fig. 10

If $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ have the same direction, the angle between them is defined to be 0, and if they have opposite directions, the angle between them is defined to be π. In any other case the angle between $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ will be between 0 and π. Thus, the angle θ between any two non-zero vectors satisfies the condition that

$0 \leq \theta \leq \pi$.

So far, we have seen how to obtain the angle between vectors by using the geometrical representation of vectors. Can we also obtain it if we use the ordered pair (or triple) representation of vectors? To answer this we define the scalar product of two vectors.

The scalar product of two vectors is a scalar.

Definition: The **scalar product** (or **dot product**, or inner product) of the two vectors $u = (a_1, a_2)$ and $v = (b_1, b_2)$ is defined to be the real number $a_1b_1 + a_2b_2$. It is denoted by $u \cdot v$. Thus,

$u \cdot v = a_1b_1 + a_2b_2$.

Similarly, the **scalar product** of $u = (a_1, a_2, a_3)$ and $v = (b_1, b_2, b_3)$ is

$u \cdot v = a_1b_1 + a_2b_2 + a_3b_3$.

Remark: Since the dot product of two vectors is a scalar, we call it the scalar product. Note that the scalar product is **not a binary operation** on R^2 or R^3. However, it has certain useful properties, some of which we give in the following theorem

Theorem 1: If $u, v, w \in R^3$ (or R^2) and $\alpha \in R$, then

a) $u \cdot u = |u|^2$, so that $u \cdot u \geq 0 \; \forall \; u$

b) $u \cdot u = 0$ iff $u = 0$

c) $u \cdot v = v \cdot u$

d) $u \cdot (v+w) = u \cdot v + u \cdot w$

e) $(\alpha u) \cdot v = \alpha (u \cdot v) = u \cdot (\alpha v)$

Proof: We shall give the proof for R^3. (You can do the proofs for R^2 similarly.)

Let $u = (a_1, a_2, a_3)$, $v = (b_1, b_2, b_3)$ and $w = (c_1, c_2, c_3)$. Then,

a) $u \cdot u = a_1^2 + a_2^2 + a_3^2 = |u|^2$.

b) $u = 0 \Rightarrow u \cdot u = 0$, since $a_1 = 0, a_2 = 0, a_3 = 0$. Conversely,
$u \cdot u = 0 \Rightarrow a_1^2 + a_2^2 + a_3^2 = 0 \Rightarrow a_1 = 0, a_2 = 0, a_3 = 0$, since the sum of non-negative real numbers is zero if and only if each one of them is zero.

c) $u \cdot v = a_1b_1 + a_2b_2 + a_3b_3 = b_1a_1 + b_2a_2 + b_3a_3 = v \cdot u$
Why don't you try finishing the proof of the theorem now? That's what we say in **E7**.

E E7) Prove (d) and (e) of Theorem 1.

Now we are in a position to obtain the angle between two vectors algebraically. We have the following theorem.

Theorem 2: If $u = (a_1, a_2, a_3)$ and $v = (b_1, b_2, b_3)$ are non-zero space vectors, and if θ is the angle between them, then $|u|\,|v| \cos \theta = u \cdot v$, that is,
$\theta = \cos^{-1} (u \cdot v / |u|\,|v|)$.

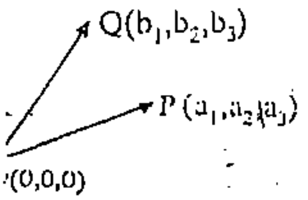


Fig. 11

Proof: Let $u = \overrightarrow{OP}$ and $v = \overrightarrow{OQ}$. So the coordinates of P and Q are (a_1, a_2, a_3) and (b_1, b_2, b_3).

First suppose $\overrightarrow{OP}$, $\overrightarrow{OQ}$ are not parallel (see Fig. 11).

By the cosine rule applied to ΔPOQ (in the plane determined by $\overrightarrow{OP}$ and $\overrightarrow{OQ}$),

$PQ^2 = OP^2 + OQ^2 - 2OP \cdot OQ \cos \theta$, i.e.,

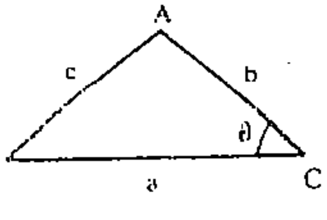
$2OP \cdot OQ \cos\theta = OP^2 + OQ^2 - PQ^2$

$= (a_1^2 + a_2^2 + a_3^2) + (b_1^2 + b_2^2 + b_3^2) - \{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2\}$.

$\therefore 2 |\overrightarrow{OP}|\,|\overrightarrow{OQ}| \cos \theta = 2(a_1b_1 + a_2b_2 + a_3b_3) = 2\, u \cdot v$ (because $OP = |\overrightarrow{OP}|$ and $OQ = |\overrightarrow{OQ}|$).

Thus, $|u|\,|v| \cos \theta = u \cdot v$.

The Cosine rule says that, for Δ ABC below, $c^2 = (a^2 + b^2 - 2ab \cos\theta)$



So we have proved Theorem 2 in the case when u and v are not parallel.

If u and v are parallel, then $v = \alpha u$ for some $\alpha \in R$ (see Sec. 2.3.2). Now, we have two possibilities : $\alpha > 0$ and $\alpha < 0$.

If $\alpha > 0$, then $\theta = 0$ and $\cos \theta = 1$. So,

$|u|\,|v| \cos \theta =$ [illegible] $u \cdot \alpha u = u \cdot v$

If $\alpha < 0$, then $|\alpha| = -\alpha$ and $\cos\theta = -1$. Hence, $|u|\,|v| \cos\theta = -|u|\,|v| = -|u|\,|\alpha u| = -|\alpha|\,|u|^2 = \alpha\,(u\cdot u) = u\cdot \alpha u = u\cdot v$.

Thus, the theorem is true in these two cases also, and hence, is true for all non-zero vectors u and v.

E8) Prove that $|u|\,|v| \cos\theta = u\cdot v$ for any two plane vectors u and v, where θ is the angle between them.

Let us look at some examples now.

Example 2: Find the angle θ between the vectors $u = (2,0)$ and $v = (1,1)$.

Solution: θ satisfies $0 \leq \theta \leq \pi$ and

$$\cos\theta = \frac{u.v}{|u|\,|v|} = \frac{(2\times 1 + 0\times 1)}{\sqrt{2^2+0^2}\ \sqrt{1^2+1^2}}$$

$$= \frac{2}{2\sqrt{2}}$$

$$= \frac{1}{\sqrt{2}},$$

so that $\theta = \pi/4$.

Example 3: Prove that the vector $v = (1/\sqrt{5}, 2/\sqrt{5})$ is equally inclined to $u = (1,0)$ and to $w = (-3/5, 4/5)$.

Solution: Note that $|u| = 1 = |v| = |w|$.
If the angles between u and v and between v and w be α and β, respectively, then

$$\cos\alpha = \frac{u\cdot v}{|u|\,|v|} = u\cdot v = 1/\sqrt{5}$$

and $\cos\beta = v\cdot w = -3/5\sqrt{5} + 8/5\sqrt{5} = 1/\sqrt{5}$

Since $0 \leq \alpha \leq \pi$, $0 \leq \beta \leq \pi$ and $\cos\alpha = \cos\beta$, we get $\alpha = \beta$.

E9) Prove that the vectors $u = (1,2,3)$ and $v = (3,0,-1)$ are perpendicular.

E E10) If the vectors u and v in each of the following are perpendicular, find a.

a) u = (1,a,2), v = (−1,2,1)

b) u = (2,−5,6), v = (1,4,a)

c) u = (a,2,−1), v = (3,a,5)

E E11) Prove that the angle between (1,0) and (−3,4) is twice the angle between (1,0) and $(1/\sqrt{5}, 2/\sqrt{5})$.

We go on to prove another property of the dot product that is very often used in the study of inner product spaces (which you will read more about in Block 4). This result is called the Schwarz Inequality.

Theorem 3: For any two vectors u,v of R^3 (or R^2), we have $|u\cdot v| \le |u|\,|v|$.

Proof: If either u = 0 or v = 0, then both sides are zero and the inequality is true. So suppose u ≠ 0 and v ≠ 0. Let θ be the angle between u and v. Then, by Theorem 2,

$\cos\theta = \frac{u\cdot v}{|u|\,|v|}$. This implies that

$|\cos\theta| = \frac{|u\cdot v|}{|u|\,|v|}$. But $|\cos\theta| \le 1$.

Thus, $\frac{|u\cdot v|}{|u|\,|v|} \le 1$, that is,

$$|u\cdot v| \le |u|\,|v|$$

Note: $|u\cdot v| = |u|\,|v|$ holds if either

i) u or v is the zero vector, or

ii) $|\cos\theta| = 1$, i.e., if θ = 0 or π.

So the two sides in Schwarz inequality are equal for non-zero vectors u and v if the vectors have the same or opposite directions.

In the next section we will see how we can use the dot product to write any vector as a linear combination of some mutually perpendicular vectors.

2.5 ORTHONORMAL BASIS

We have seen how to calculate the angle between any two vectors. If the angle between two non-zero vectors u and v is $\pi/2$ then they are said to be **orthogonal**. That is, if u and v are mutually perpendicular then they are orthogonal. Now, if u and v are orthogonal, then, by Theorem 2,

$$0 = \cos \pi/2 = \frac{u\cdot v}{|u|\,|v|} \Rightarrow u\cdot v = 0.$$

Conversely, if u, v are non-zero and if $u\cdot v = 0$, then the angle θ between them satisfies

$$\cos\theta = \frac{u\cdot v}{|u|\,|v|} = 0, \text{ so that } \theta = \pi/2.$$

Thus, for non-zero vectors u and v, $u\cdot v = 0$ iff u and v are orthogonal.

An important set of orthogonal vectors in R^2 is {i,j} (see Fig. 12(a)), where i = (1,0) and j = (0,1). Thus, i and j are unit vectors along the x and y axes, respectively. They are orthogonal because $i\cdot j = 1\cdot 0 + 0\cdot 1 = 0$.

Similarly, in R^3, i = (1,0,0), j = (0,1,0), k = (0,0,1), are mutually orthogonal (see Fig. 12(b)), since

$i\cdot j = 1\cdot 0 + 0\cdot 1 + 0\cdot 0 = 0$, $j\cdot k = 0\cdot 0 + 1\cdot 0 + 0.1 = 0$ and $k\cdot i = (0\cdot 1 + 0\cdot 0 + 1\cdot 0) = 0$.

The vectors a,b,c,... are called mutually orthogonal if each of them is orthogonal to each of the others.

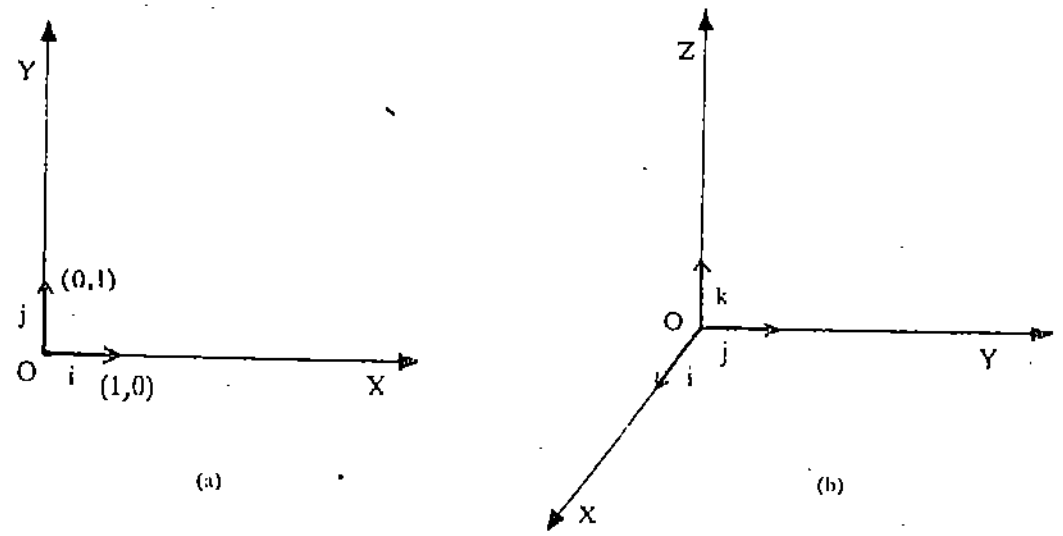


Fig. 12

Note that, I and j in R^2, and i,j,k in R^3, are not only mutually orthogonal, but each of them is also a unit vector. Such a set of vectors is called an orthonormal system.

Definition: A set of vectors of R^3 (or R^2) are said to form an **orthonormal system** if each vector in the set is a unit vector and any two vectors of the set are mutually orthogonal.

An orthonormal system is very important because every vector in R^3 (or R^2) can be expressed as a linear combination of the vectors in such a system. In the following theorem we will prove that any vector in R^3 is a linear combination of the orthogonal system {i,j,k}.

Theorem 4: Every vector in R^3 is a linear combination of **i,j,k**.

Proof: Let $x = (x_1,x_2,x_3)$ be any space vector. Then

$$x = (x_1,x_2,x_3) = x_1(1,0,0) + x_2(0,1,0) + x_3(0,0,1)$$
$$= x_1 i + x_2 j + x_3 k.$$

Thus, our theorem is proved.

Note: In the proof above, $x_1 = x\cdot i$, $x_2 = x\cdot j$ and $x_3 = x\cdot k$.

In fact, if {u,v,w} is any orthonormal system in R^3, then every space vector x can be expressed as a linear combination of u,v,w as

$$x = (x\cdot u)u + (x\cdot v)v + (x\cdot w)w.$$

Since the proof of this is a little complicated we will not give it over here.

Remark: The result given in Theorem 4 also holds good for R^2, if we replace {i,j,k} by {i = (1,0), j = (0,1)}. It is also true that every vector in R^2 can be written as a linear combination of an orthonormal system {u,v} in R^2.

Since three orthonormal vectors in R^3 have the property that all vectors in R^3 can be written in terms of these, we say that these vectors form an **orthonormal basis** for the vector space R^3. (We explain the term 'basis' later, in Unit 4.) Similarly, two orthonormal vectors in R^2 form an orthonormal basis of R^2.

Example 4: Prove that

$u = (1/\sqrt{3})(i - j + k)$

$v = (1/\sqrt{6})(2i + j - k)$, and

$w = (1/\sqrt{2})(j+k)$

form an orthonormal basis of R^3. Express $x = -i+3j+4k$ as a linear combination of u,v,w.

Solution: Since $|u| = 1 = |v| = |w|$, and $u\cdot v = u\cdot w = w\cdot v = 0$, we see that {u,v,w} is an orthonormal system in R^3. Therefore, it forms an orthonormal basis of R^3. Thus, from what you have just read, you know that x can be written as

$(x\cdot u)u + (x\cdot v)v + (x\cdot w)w$. Now

$$x\cdot u = (1/\sqrt{3})(-i+3j+4k)\cdot(i - j + k)$$
$$= (1/\sqrt{3})\ (-i\cdot i - 3j\cdot j + 4k\cdot k)$$
$$= (1/\sqrt{3})\ (-1 - 3 + 4) = 0$$

Next,

$$x\cdot v = (1/\sqrt{6})(-i + 3j+4k).(2i+j-k)$$
$$= (1/\sqrt{6})(-2+3-4) = -3/\sqrt{6}, \text{ and}$$
$$x\cdot w = (1/\sqrt{2})(-i + 3j + 4k).(j + k)$$
$$= (1/\sqrt{2})(3 + 4) = 7/\sqrt{2}.$$

Hence,

$$x = (-3/\sqrt{6})\,v + (7/\sqrt{2})\,w$$

E E12) If $x = 3i - j - k$, express x as a linear combination of u,v,w of the example above.

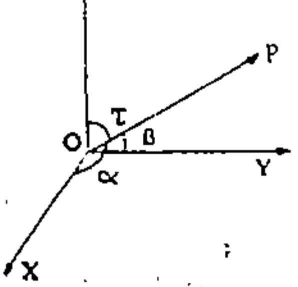


Fig. 13

Let us now see how to find the angle that a space vector makes with each of the axes. You know that, for any vector x in R^3, $x = (x\cdot i)i + (x\cdot j)j + (x\cdot k)k$. Also, i,j,k lie along the x,y and z axes, respectively. Suppose x makes angles of α, β, τ with the x,y and z axes respectively (see Fig. 13). Then, by Theorem 2,

$$\cos\alpha = \frac{x\cdot i}{|x|\,|i|} = \frac{x\cdot i}{|x|}$$

Similarly, $\cos\beta = \frac{x \cdot j}{|x|}$ and $\cos\tau = \frac{x \cdot k}{|x|}$. These quantities are called the direction cosines of x. Thus, the cosines of the angles formed by $x = \overrightarrow{OP}$ with the positive directions of the three axes are its **direction cosines.**

We have just seen that:

If x is a non-zero vector in R^3, then its direction cosines are

$$\frac{x \cdot i}{|x|}, \quad \frac{x \cdot j}{|x|}, \quad \frac{x \cdot k}{|x|}$$

For example, the direction cosines of i are 1,0,0 because

$$\frac{i \cdot i}{|i|} = 1, \quad \frac{i \cdot j}{|i|} = 0, \quad \frac{i \cdot k}{|i|} = 0.$$

Similarly, the direction cosines of $u = (a_1,a_2,a_3)$ are

$$\frac{a_1}{\sqrt{a_1^2+a_2^2+a_3^2}}, \quad \frac{a_2}{\sqrt{a_1^2+a_2^2+a_3^2}} \quad \text{and} \quad \frac{a_3}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

This is because,

$$\frac{u \cdot i}{|u|} = \frac{(a_1,a_2,a_3)\cdot(1,0,0)}{\sqrt{a_1^2+a_2^2+a_3^2}} = \frac{a_1}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

Similarly,

$$\frac{u \cdot j}{|u|} = \frac{a_2}{\sqrt{a_1^2+a_2^2+a_3^2}}, \quad \frac{u \cdot k}{|u|} = \frac{a_3}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

E E13) Find the direction cosines of the vector i + j.

We now give a very nice property pertaining to direction cosines.

Theorem 5: If $\cos\alpha$, $\cos\beta$, $\cos\tau$ are the direction cosines of a non-zero vector u, then $\cos^2\alpha + \cos^2\beta + \cos^2\tau = 1$

Proof: You have just seen that the direction cosines of $u = (a_1,a_2,a_3)$ are

$$\cos\alpha = \frac{a_1}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

$$\cos\beta = \frac{a_2}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

$$\cos\tau = \frac{a_3}{\sqrt{a_1^2+a_2^2+a_3^2}}$$

from which it is obvious that

$$\cos^2\alpha + \cos^2\beta + \cos^2\tau = 1$$

This theorem ends our discussion of the scalar product of vectors.

Before going further, we mention another kind of product of two vectors in R^3, namely, the cross product. The cross product of two vectors a and b in R^3, denoted by $a \times b$, is defined to be the vector whose direction is perpendicular to the plane of a and b, and magnitude is |a| |b| sin θ, where θ is the angle between a and b (see Fig. 14).

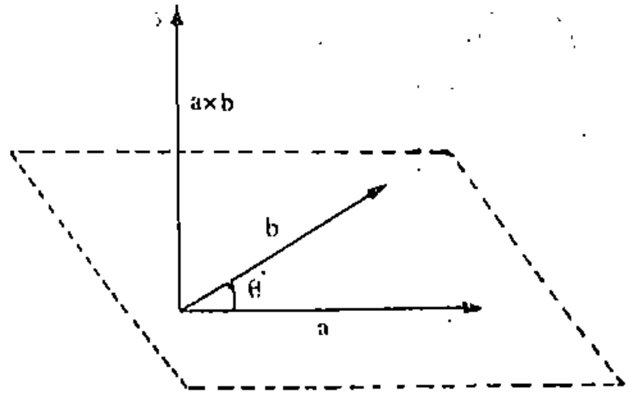


Fig. 14

Thus, $\mathbf{a} \times \mathbf{b} = (|\mathbf{a}|\ |\mathbf{b}| \sin\theta)\,\mathbf{n}$, where n is a unit vector perpendicular to the plane of a and b.

For example, $\mathbf{i} \times \mathbf{j} = \mathbf{k}$, $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{i} = \mathbf{j}$.

Note that this way of multiplying two vectors is not possible in $\mathbf{R}^2$

Now let us try to represent some geometrical concepts by using vectors.

2.6 VECTORS AND GEOMETRY OF SPACE

In this section we will obtain the equations of a line, a plane and a sphere in terms of vectors.

2.6.1 Vector Equation of a Line

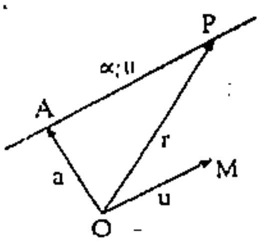


Fig. 15

The position vector of a point P is $\overrightarrow{OP}$

Let A be a point in $\mathbf{R}^3$ and $\overrightarrow{OA}$ be denoted by a. Let u be a given vector in $\mathbf{R}^3$. Then the equation of the line through A and parallel to u is

$\mathbf{r} = \mathbf{a} + \alpha\mathbf{u}$,

where α is a real parameter.

This means that the position vector r, of any point P on such a line, satisfies $\mathbf{r} = \mathbf{a} + \alpha\mathbf{u}$, for some real number α. Conversely, for every real number α, the point whose position vector is $\mathbf{a} + \alpha\mathbf{u}$ is on this line (see Fig. 15).

The vector equation $\mathbf{r} = \mathbf{a} + \alpha\mathbf{u}$, of a line, corresponds to the Cartesian equation

$$\frac{x - x_0}{l} = \frac{y - y_0}{m} = \frac{z - z_0}{n}.$$

Example 5: Find the equation of the line through the point A(1, −1, 1) and parallel to the line joining B(1,2,3) and C(−2,0,1).

Solution: The position vector a of A is (1, −1, 1).

$$\text{Also } \overrightarrow{BC} = \overrightarrow{OC} - \overrightarrow{OB}$$
$$= (-2,0,1) - (1,2,3)$$
$$= (-3, -2, -2)$$

Hence, $\mathbf{u} = (-3, -2, -2)$

Thus, the vector equation of the line through A and parallel to BC is

$$\mathbf{r} = \mathbf{a} + \alpha\mathbf{u} = (1,-1,1) + \alpha(-3,-2,-2)$$
$$= (1,-1,1) + (-3\alpha, -2\alpha, -2\alpha)$$
$$= (-3\alpha + 1, -2\alpha - 1, -2\alpha + 1)$$

Remark: Whatever has been discussed above is also true for R^2. That is, the equation of any line in R^2 that passes through $a = (a_1,a_2)$ and is parallel to a given vector $u = (u_1,u_2)$ is $r = a + \alpha u$, $\alpha \in R$.

This corresponds to the Cartesian equation $\frac{x - x_0}{l} = \frac{y - y_0}{m}$.

E14) Find the vector equation of the line passing through a = (1,0), and parallel to the y-axis.

Now how do we get the vector equation of a straight line in R^3, which passes through points A and B, whose position vectors are a and b, respectively?

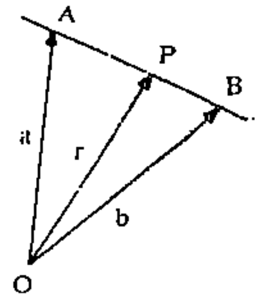


Fig. 16.

Since $\overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} = b - a$ (see Fig. 16), we want the equation of a line passing through A and parallel to the vector b−a.
Hence the desired equation is

$r = a + \alpha (b - a)$.

This equation corresponds to the Cartesian equation

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} = \frac{z - z_1}{z_2 - z_1}$$

of the line passing through (x_1,y_1,z_1) and (x_2,y_2,z_2).

Remark: The vector equation of any line in R^2 passing through $a = (a_1,a_2)$ and $b = (b_1,b_2)$ is $r = a + \alpha (b - a)$.

Example 6: What is the vector equation of a line passing through j = (0,1,0) and k = (0,0,1)?

Solution: Now j − k = (0,1,0) − (0,0,1) = (0,1,−1).
Thus, the required equation is
$r = k + \alpha(j - k) = (0,0,1) + \alpha(0,1,-1)$.
$= (0, \alpha, 1 - \alpha)$

E15) Find the vector equation of the line passing through i and i + j + k. What are the direction cosines of the vector on this line which corresponds to the value $\alpha = 1$?

Now let us see how to obtain the equation of a plane in terms of vectors.

2.6.2 Vector Equation of a Plane

Let A,B,C be non-collinear points in R^3 with position vectors a,b,c, respectively. Then, from Euclidean geometry you know that the three points A,B,C determine a unique plane. The **vector equation of the plane** determined by A,B,C is
$r = a + \alpha(b-a) + \mu (c-a)$, where α, μ are any real numbers. Why is this the equation?

μ is the Greek letter mu

Well, suppose you take any point P in the plane determined by A, B and C. Then, since A, B and C are not collinear, the vector $\overrightarrow{AP}$ is a linear combination of the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ (see Fig. 17). That is, $\overrightarrow{AP} = \alpha \overrightarrow{AB} + \mu \overrightarrow{AC}$, $\alpha, \mu \in \mathbf{R}$. Now, $\overrightarrow{OP} = \overrightarrow{OA} + \overrightarrow{AP} =$ $\mathbf{a} + \alpha\overrightarrow{AB} + \mu \overrightarrow{AC} = \mathbf{a} + \alpha(\mathbf{b}-\mathbf{a}) + \mu(\mathbf{c}-\mathbf{a})$.

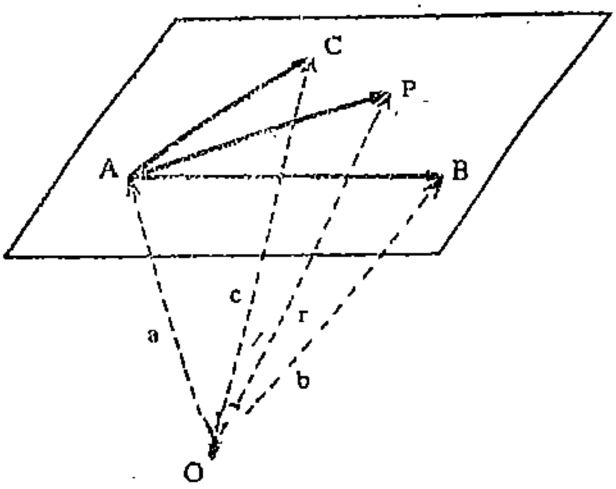


Fig. 17

We can rewrite the equation of the plane containing the points A, B, C as

$\mathbf{r} = (1 - \alpha - \mu)\mathbf{a} + \alpha \mathbf{b} + \mu \mathbf{c}$,

This shows us that **r** is a linear combination of the vectors **a**, **b** and **c**.

Example 7: Find the vector equation of the plane determined by the points (0,1,1), (2,1,−3) and (1,3,2). Also find the point where the line $\mathbf{r} = (1+2\alpha)\,\mathbf{i} + (2-3\alpha)\mathbf{j} - (3+5\alpha)\mathbf{k}$ intersects this plane.

Solution: The position vectors of the three given points are

$\mathbf{j} + \mathbf{k},\ 2\mathbf{i} + \mathbf{j} - 3\mathbf{k},\ \mathbf{i} + 3\mathbf{j} + 2\mathbf{k}$.

Therefore, the equation of the plane is

$\mathbf{r} = \mathbf{j} + \mathbf{k} + s(2\mathbf{i} - 4\mathbf{k}) + t\,(\mathbf{i} + 2\mathbf{j} + \mathbf{k})$, that is,

$\mathbf{r} = (2s + t)\mathbf{i} + (1 + 2t)\,\mathbf{j} + (1-4s + t)\,\mathbf{k}$, where s, t are real parameters.

The second part of the question requires us to find the point of intersection of the given line and the plane. This point must satisfy the equations of the plane and this line. Thus, s, t and α must satisfy

$2s + t = 1 + 2\alpha,\ 1 + 2t = 2 - 3\alpha,\ 1-4s + t = -3-5\alpha.$

When these simultaneous equations are solved, we get $s = 2, t = -1, \alpha = 1$. Putting this value of α in the equation of the line, we find the position vector **r**, of the point of intersection is

$\mathbf{r} = 3\mathbf{i} - \mathbf{j} - 8\mathbf{k}$,

so that the required point is (3, −1, −8).

E E16) Find the equation of the plane passing through **i**, **j** and **k**.

We will now give the vector equation of a plane when we know that it is perpendicular to a fixed unit vector **n**, and we know the distance of the origin from it is d.

The required equation is

$\mathbf{r} \cdot \mathbf{n} = d$

Note that $d \geq 0$ always, being the distance from the origin.

The equation $\mathbf{r}\cdot\mathbf{n} = d$ corresponds to the Cartesian equation $ax + by + cz = d$, of a plane.

Example 8: Find the direction cosines of the perpendicular from the origin to the plane $\mathbf{r}\cdot(6\mathbf{i} - 3\mathbf{j} - 2\mathbf{k}) + 1 = 0$.

Solution: We rewrite the given equation as

$\mathbf{r}\cdot(6\mathbf{i} - 3\mathbf{j} - 2\mathbf{k}) = -1$

Now $|6\mathbf{i} - 3\mathbf{j} - 2\mathbf{k}| = \sqrt{36 + 9 + 4} = 7$. Thus,

$\left|\frac{6}{7}\mathbf{i} - \frac{3}{7}\mathbf{j} - \frac{2}{7}\mathbf{k}\right| = 1$ and $\frac{6}{7}\mathbf{i} - \frac{3}{7}\mathbf{j} - \frac{2}{7}\mathbf{k}$ is a unit vector. Then

$\mathbf{r}\cdot\left(\frac{-6}{7}\mathbf{i} + \frac{3}{7}\mathbf{j} + \frac{2}{7}\mathbf{k}\right) = \frac{1}{7}$

is the equation of the given plane, in the form $\mathbf{r}\cdot\mathbf{n} = d$, with $d \geq 0$ and $\mathbf{n}$ being a unit vector. This shows that the perpendicular unit vector from the origin to the plane is

$\mathbf{n} = \frac{-6}{7}\mathbf{i} + \frac{3}{7}\mathbf{j} + \frac{2}{7}\mathbf{k}$. Its direction cosines are what we want.

They are $-\frac{6}{7}, \frac{3}{7}, \frac{2}{7}$.

E E17) What is the distance of the origin from the plane
$\mathbf{r}\cdot(\mathbf{i}+\mathbf{j}+\mathbf{k}) + 5 = 0$?

Let us now look at the vector equation of a sphere.

2.6.3 Vector Equation of a Sphere

As you know, a sphere is the locus of a point in space which is at a constant distance from a fixed point. The constant distance is called the **radius** and the fixed point is called the **centre** of the sphere. If the radius is a and the centre is (c_1, c_2, c_3), then the Cartesian equation of the sphere is

$$(x-c_1)^2 + (y-c_2)^2 + (z-c_3)^2 = a^2.$$

The vector equation of the same sphere (see Fig. 18) is $|\mathbf{r} - \mathbf{c}| = a$, where $\mathbf{c} = (c_1, c_2, c_3)$.

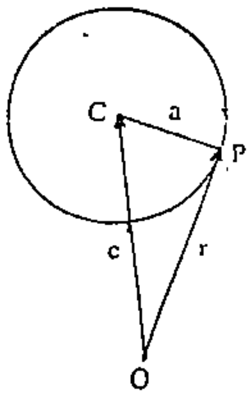


Fig. 18

In particular, the vector equation of a sphere whose centre is the origin and radius is a is $|\mathbf{r}| = a$.

We give the following example.

Example 9: Find the radius of the circular section of the sphere $|\mathbf{r}| = 5$ by the plane $\mathbf{r}\cdot(\mathbf{i}+\mathbf{j}+\mathbf{k}) = 3\sqrt{3}$.

Solution: The sphere $|\mathbf{r}| = 5$ has centre the origin, and radius 5. The plane $\mathbf{r}\cdot(\mathbf{i}+\mathbf{j}+\mathbf{k}) = 3\sqrt{3}$ can be rewritten as $\mathbf{r}\cdot(1/\sqrt{3})(\mathbf{i}+\mathbf{j}+\mathbf{k}) = 3$, in which $(1/\sqrt{3})(\mathbf{i}+\mathbf{j}+\mathbf{k})$ is a unit vector. This shows that the distance of this plane from the origin is 3. So the plane and the sphere intersect, giving a circular section of the sphere. In Fig. 19 OP = 5, ON = 3. Hence, $NP^2 = OP^2 - ON^2 = 5^2 - 3^2 = 4^2$. So, the required radius, NP = 4.

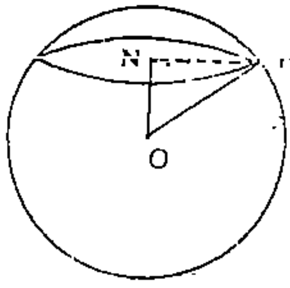


Fig. 19

E18) Find the radius of the circular section of the sphere $|r| = 13$ by the plane $r.(2i+3j+6k) = 35$.

Let us finally recapitulate what we have done in this unit.

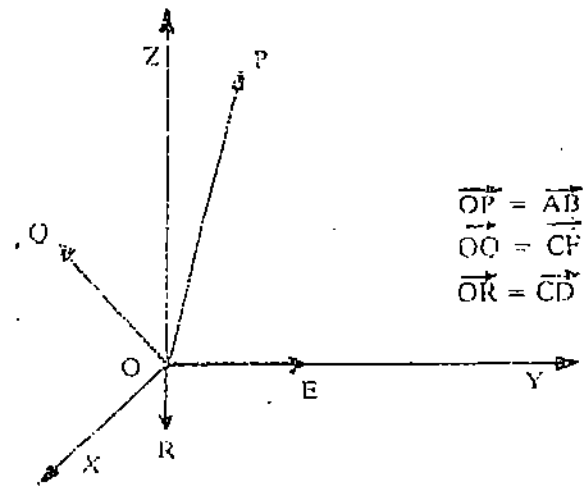
2.7 SUMMARY

We end this unit with summarising what we have covered in it. We have

1) defined vectors as directed line segments, and as ordered pairs or triples.
2) introduced you to the operations of vector addition and scalar multiplication in $\mathbf{R}^2$ and $\mathbf{R}^3$.
3) defined the scalar products of vectors, and used this concept for obtaining direction cosines of vectors.
4) given the vector equations of a line, a plane and a sphere.

2.8 SOLUTIONS/ANSWERS

E1)



E2) a) 2,0 b) false c) 3

E3) The proof is the same as that for $\mathbf{R}^3$, except that you will deal with ordered pairs instead of ordered triples.

E4) Let $u = (a,b,c)$. Then $|\alpha u| = \sqrt{\alpha^2 a^2 + \alpha^2 b^2 + \alpha^2 c^2} = |\alpha|\,|u|$.

E5) The proof is the same as that for $\mathbf{R}^2$, except that you will deal with triples, instead of pairs.

E6) Let (a,b) be any plane vector. Then $(a,b) = a(1,0) + b(0,1)$, and hence, is a linear combination of (1,0) and (0,1).

E7) d) $u.(v+w) = (a_1,a_2,a_3).(b_1+c_1, b_2+c_2, b_3+c_3)$

$= a_1(b_1+c_1) + a_2(b_2+c_2) + a_3(b_3+c_3)$

$= (a_1b_1+a_2b_2+a_3b_3) + (a_1c_1+a_2c_2+a_3c_3)$

$= u.v + u.w$

e) $(\alpha u).v = (\alpha a_1, \alpha a_2, \alpha a_3) . (b_1, b_2, b_3)$
$= \alpha a_1 b_1 + \alpha a_2 b_2 + \alpha a_3 b_3$
$= \alpha (u.v)$

You can similarly show that $(\alpha u).v = u.(\alpha v)$.

E8) First consider any two vectors u and v, which are not in the same line. Let $u = (a_1, a_2)$ and $v = (b_1, b_2)$. Then, as in Theorem 2, $|u|\,|v| \cos\theta = u.v$. Next, consider the case when u and v are in the same line. Then $u = \alpha v$, for $\alpha \in R$. Then, as in Theorem 2, you can again prove that $|u|\,|v| \cos\theta = u.v$.

E9) Suppose θ is the angle between them. Then

$\cos\theta = \frac{u.v}{|u|\,|v|} = 0$. Also $0 \le \theta \le \pi$. This gives $\theta = \pi/2$, that is u and v are perpendicular.

E10) Now u and v are perpendicular iff $u.v = 0$.
a) $u.v = 0 \Rightarrow 1 \times (-1) + a \times 2 + 2 \times 1 = 0 \Rightarrow 1 + 2a = 0$
$\Rightarrow a = \frac{-1}{2}$
b) $a = 3$, c) $a = 1$

E11) Let $u = (1,0)$, $v = (-3,4)$, $w = (1/\sqrt{5}, 2/\sqrt{5})$. Let the angles between u and v and u and w be α and β, respectively.

Then we have to show that $\alpha = 2\beta$. Now, $\cos\alpha = \frac{u.v}{|u|\,|v|} = \frac{-3}{5}$

and $\cos\beta = \frac{u.w}{|u|\,|w|} = \frac{1}{\sqrt{5}}$

A result from trigonometry is
$\cos 2\theta = 2\cos^2\theta - 1$, for any angle θ.

Therefore, $\cos 2\beta = 2(1/5) - 1 = -3/5 = \cos\alpha$. Since $\cos\beta$ is positive, $0 < \beta < \pi/2$.

Therefore, $0 < 2\beta < \pi$. Also $0 < \alpha < \pi$, and $\cos 2\beta = \cos\alpha$. Hence, $2\beta = \alpha$.

E12) $x = (x.u)u + (x.v)v + (x.w)w$
Now, $x.u = (1/\sqrt{3})(3i - j - k).(i - j + k) = 1/\sqrt{3}\,(3 + 1 - 1) = \sqrt{3}$.
$x.v = \sqrt{6}$ and $x \cdot w = -\sqrt{2}$.
Therefore, $x = \sqrt{3}\, u + \sqrt{6}\, v - \sqrt{2}\, w$.

E13) Since $|i+j| = \sqrt{2}$, we get the direction cosines to be $1/\sqrt{2}, 1/\sqrt{2}, 0$.

E14) $j = (0,1)$ is a vector along the y-axis. Thus, our line should be parallel to j. Therefore, the required equation is $r = a + \alpha j = (1,0) + \alpha(0,1) = (1,\alpha)$. $\alpha \in R$.

E15) The required equation is $r = i + \alpha(i+j+k-i) = i + \alpha(j+k)$
$= (1,0,0) + \alpha(0,1,1) = (1,\alpha,\alpha)$, $\alpha \in R$.
When $\alpha = 1$, we get the vector $(1,1,1)$. Its direction cosines are $1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}$.

E16) The required equation is $r = i + s(j-i) + t(k-i)$, where $s, t \in R$. This gives us $r = (1,0,0) + s(-1,1,0) + t(-1,0,1) = (1-s-t, s, t)$.

E17) First we put the equation of the plane in the form $r.n = d$, where n is a unit vector and $d \ge 0$. Now $|i+j+k| = \sqrt{3}$. Therefore, $|1/\sqrt{3}(i+j+k)| = 1$, and hence, $1/\sqrt{3}(i+j+k)$ is a unit vector. $\therefore$, $(-1/\sqrt{3})(i+j+k)$ is also a unit vector. Now the given plane's equation is $r \cdot (i+j+k) = -5$
$\Rightarrow r.(-1/\sqrt{3})(i+j+k) = 5/\sqrt{3}$.
Thus, the required distance is $d = 5/\sqrt{3}$.

E18) The centre of the sphere is $(0,0,0)$, and radius is 13. The given plane is $r.(2/7\, i + 3/7\, j + 6/7\, k) = 5$, in the form $r.n = d$. Therefore, the radius of the circular section is $\sqrt{13^2 - 5^2} = 12$.

UNIT 3 VECTOR SPACES

Structure

3.1 INTRODUCTION

In this unit we begin the study of vector spaces and their properties. The concepts that we will discuss here are very important, since they form the core of the rest of the course. In Unit 2 we studied R^2 and R^3. We also defined the two operations of vector addition and scalar multiplication on them along with certain properties. This can be done in a more general setting. That is, we may start with any set V (in place of R^2 or R^3) and convert V into a vector space by introducing "addition" and "scalar multiplication" in such a way that they have all the basic properties which vector addition and scalar multiplication have in R^2 and R^3. We will prove a number of results about the general vector space V. These results will be true for all vector spaces — no matter what the elements are. To illustrate the wide applicability of our results, we shall also give several examples of specific vector spaces.

We shall also study subsets of a vector space which are vector spaces themselves. They are called subspaces. Finally, using subspaces, we will obtain new vector spaces from given ones.

Since this unit forms part of the backbone of the course, be sure that you understand each concept in it.

Objectives

After studying this unit, you should be able to

- define and recognise a vector space;
- give a wide variety of examples of vector spaces;
- determine whether a given subset of a vector space is a subspace or not;
- explain what the linear span of a subset of a vector space is;
- differentiate between the sum and the direct sum of subspaces;
- define and give examples of cosets and quotient spaces.

3.2 WHAT ARE VECTOR SPACES?

You have already come across the algebraic structure called a field in Unit 1. We now build another algebraic structure from a set, by defining on it the operations of addition and multiplication by elements of a field. This is a vector space. We give the definition of a vector space now. As you read through it you can keep in mind the example of the vector space R^2 over R (Unit 2).

Definition: A set V is called a vector space over a field F if it has two operations, namely addition (denoted by +) and multiplication of elements of V by elements of F (denoted by ·), such that the following properties hold:

VS1) + is a binary operation, i.e., $u + v \in V \; \forall \; u, v \in V$.

VS2) + is associative, i.e., $(u + v) + w = u + (v + w) \; \forall \; u, v, w \in V$.

VS3) V has an identity element with respect to +, i.e.,
$\exists \; 0 \in V$ such that $0 + v = v = v + 0 \; \forall \; v \in V$

VS4) Every element of V has an inverse with respect to +: For every $u \in V$, $\exists \; v \in V$ such that $u + v = 0$. v is called the additive inverse of u, and is written as $-u$.

VS5) + is commutative, i.e., $u + v = v + u \; \forall \; u, v \in V$.

VS6) $\cdot : F \times V \to V : \cdot(\alpha, v) = \alpha.v$ is a well defined operation, i.e., $\forall \; \alpha \in F$ and $v \in V$, $\alpha.v \in V$.

VS7) $\forall \; \alpha \in F$ and $u, v \in V$, $\alpha.(u+v) = \alpha.u + \alpha.v$.

VS8) $\forall \; \alpha, \beta \in F$ and $v \in V$, $(\alpha+\beta).v = \alpha.v + \beta.v$

VS9) $\forall \; \alpha, \beta \in F$ and $v \in V$, $(\alpha \beta).v = \alpha.(\beta.v)$

VS10) $1.v = v$, for all $v \in V$.

When V is a vector space over **R**, we also call it a **real vector space**. Similarly, if V is defined over **C**, it is also called a **complex vector space**.

The product of $\alpha \in F$ and $v \in V$, in the definition, is often denoted by αv instead of $\alpha.v$. Note that this product is a vector. This operation is called **scalar multiplication**, because the elements of F are called **scalars**. Elements of V are called **vectors**.

Now that the additive inverse of a vector is defined (in VS4), we can give another definition.

Definition: If u,v belong to a vector space V, we define their difference $u - v$ to be $u + (-v)$.
For example, in $\mathbf{R}^2$ we have $(3,5) - (1,0) = (3,5) + (-1,0) = (2,5)$.

After going through Unit 2 and the definition of a vector space it must be clear to you that $\mathbf{R}^2$ and $\mathbf{R}^3$, with vector addition and scalar multiplication, are vector spaces.

We now give some more examples of vector spaces.

Example 1: Show that **R** is a vector space over itself.

Solution: '+' is associative and commutative in **R**. The additive identity is 0 and the additive inverse of $x \in \mathbf{R}$ is $-x$. The scalar multiplication is the ordinary multiplication in **R**, and satisfies the properties VS7-VS10.

Example 2: For any positive integer n, show that the set
$\mathbf{R}^n = \{(x_1, x_2, \ldots\ldots, x_n) \mid x_i \in \mathbf{R}\}$ is a vector space over **R**, if we define vector addition and scalar multiplication as:
$(x_1, x_2, \ldots\ldots, x_n) + (y_1, y_2, \ldots\ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots\ldots, x_n + y_n)$, and
$\alpha(x_1, x_2, \ldots\ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots\ldots, \alpha x_n)$, $\alpha \in \mathbf{R}$.

For any field F,
$F^n = \{(x_1, \ldots, x_n) \mid x_i \in F\}$.
Every element of F^n is called an n-tuple of elements of F

Solution: The properties VS1 – VS10 are easily checked. Since '+' is associative and commutative in **R**, you can check that '+' is associative and commutative in $\mathbf{R}^n$ also. Further, the identity for addition is $(0, 0, \ldots\ldots, 0)$ because
$(x_1, x_2, \ldots\ldots, x_n) + (0, 0, \ldots\ldots, 0)$
$= (x_1 + 0, x_2 + 0, \ldots\ldots, x_n + 0) = (x_1, x_2, \ldots\ldots, x_n)$.

The additive inverse of $(x_1, \ldots\ldots, x_n)$ is $(-x_1, \ldots\ldots, -x_n)$.

For $\alpha, \beta \in \mathbf{R}$, $(\alpha + \beta)(x_1, \ldots\ldots, x_n) = ((\alpha + \beta) x_1, \ldots\ldots, (\alpha + \beta) x_n)$
$= (\alpha x_1 + \beta x_1, \ldots\ldots, \alpha x_n + \beta x_n)$
$= (\alpha x_1, \ldots\ldots, \alpha x_n) + (\beta x_1, \ldots\ldots, \beta x_n)$
$= \alpha(x_1, \ldots\ldots, x_n) + \beta(x_1, \ldots\ldots, x_n)$

E E1) Prove that properties VS7, VS9 and VS10 hold for $\mathbf{R}^n$.

Caution: The symbol '+' on the left hand side of

$(x_1, x_2, \ldots\ldots, x_n) + (y_1, y_2, \ldots\ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots\ldots, x_n + y_n)$

and the same symbol on the right hand side is used for different operations. On the left, we are adding two vectors so it is used for the addition operation in $\mathbf{R}^n$; on the right, we are adding only two real numbers, x_1 and y_1 or x_2 and y_2, and so on. So, on the right, + indicates the addition operation in $\mathbf{R}$.

Example 3: Consider the set C of all complex numbers. Prove that C is a real vector space.

Solution: Addition of two complex numbers has been defined in Sec. 1.5, and satisfies VS1 – VS5.

If $\alpha \in \mathbf{R}$ and $u \in \mathbf{C}$ then α and u are both elements of C, since $\mathbf{R} \subseteq \mathbf{C}$. Thus $\alpha u \in \mathbf{C}$, so that scalar multiplication is also defined. Can you see that this operation has the properties VS6 – VS10? Once you answer this question you will see that C is a real vector space.

Note: C is also a complex vector space. This can be shown on the lines of Example 1.

Example 4: Let P be the set of all polynomials in x with real coefficients, i.e.,

$$\mathbf{P} = \left\{ \sum_{i=0}^{n} a_i x^i \mid n \text{ is a positive integer, } a_i \in \mathbf{R} \right\}$$

Show that P is a real vector space.

Solution: The sum of two polynomials is a polynominal (for example, $(x^2 + 1) + (2x + 3) = x^2 + 2x + 4$.) This addition operation on P is commutative and associative; the polynomial 0 is the additive identity and, given a polynomial $p(x) = a_0 + a_1x + \ldots\ldots + a_nx^n$, its additive inverse is

$-p(x) = -a_0 - a_1x - \ldots\ldots - a_nx^n$. For $\alpha \in \mathbf{R}$ and $\sum_{i=0}^{n} a_ix^i \in \mathbf{P}$, define

$$\alpha \cdot \sum_{i=0}^{n} a_ix^i = \sum_{i=0}^{n} (\alpha a_i)x^i.$$

(For example, $3 \cdot (x^2 + 4) = 3x^2 + 3.4 = 3x^2 + 12$.)

This scalar multiplication has all the requisite properties VS6 – VS10. Thus, the set P of all polynomials is a real vector space.

Example 5: Consider the set S, of all functions from R to R. Define addition on S as follows:

If $f \in S$, $g \in S$, then $f + g$ is the function defined by

$(f + g)(x) = f(x) + g(x) \ \forall x \in \mathbf{R}$.

(This is called pointwise addition.)

Define scalar multiplication as follows:

For $\alpha \in \mathbf{R}$, $f \in S$, let αf be the function given by

$(\alpha f)(x) = \alpha \cdot f(x) \; \forall \; x \in \mathbf{R}$.

Show that S is a real vector space.

Solution: The properties VS1 – VS5 are satisfied. The additive identity is the function $0(x)$ such that $0(x) = 0$ for all $x \in \mathbf{R}$.

The inverse of f is $-f$ where $(-f)(x) = -[f(x)] \; \forall \; x \in \mathbf{R}$.

E2) Check that VS6 – VS10 hold true for the set S in the example above.

Example 6: Let $V \subseteq \mathbf{R}^2$ be given by

$V = \{(x,y) \mid x,y \in \mathbf{R} \text{ and } y = 5x\}$

We define addition and scalar multiplication on V to be the same as in $\mathbf{R}^2$, i.e.,

$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$ and

$\alpha(x,y) = (\alpha x, \alpha y)$, for $\alpha \in \mathbf{R}$.

Show that V is a real vector space.

Solution: First note that addition is a binary operation on V. This is because

$(x_1, y_1) \in V, (x_2, y_2) \in V \Rightarrow y_1 = 5x_1, y_2 = 5x_2 \Rightarrow y_1 + y_2 = 5(x_1 + x_2)$

$\Rightarrow (x_1+x_2, y_1+y_2) \in V$.

The addition is also associative and commutative, since it is so in $\mathbf{R}^2$. Next, the additive identity for $\mathbf{R}^2$, (0,0), belongs to V and is the additive identity for V. Finally, if $(x,y) \in V$ (i.e., $y = 5x$), then its additive inverse $-(x,y) = (-x, -y) \in \mathbf{R}^2$.

Also $-y = 5(-x)$. So that, $-(x,y) \in V$.

That is, $(x,y) \in V \Rightarrow -(x,y) \in V$.

Thus, VS1 – VS5 are satisfied by addition on V.

As for scalar multiplication, if $\alpha \in \mathbf{R}$ and $(x,y) \in V$, then $y = 5x$, so that $\alpha y = 5(\alpha x)$.

$\therefore$, $\alpha(x,y) \in V$.

That is, VS6 is satisfied.

The properties VS7 – VS10 also hold good, since they do so for $\mathbf{R}^2$.

Thus V becomes a real vector space.

Check your understanding of vector spaces by trying the following exercises.

E3) Let V be the subset of complex numbers given by

$V = \{x + ix \mid x \in \mathbf{R}\}$.

Show that, under the usual addition of complex numbers and scalar multiplication defined by $\alpha(x + ix) = \alpha x + i(\alpha x)$, V is a real vector space.

E4) Let $Q = \{ax^2 + bx + c | a, b, c \in C\}$, i.e.,
Q is the set of all complex polynomials (i.e., polynomials with complex coefficients) of degree at most 2. Under the usual operations of addition and scalar multiplication, prove that Q is a complex vector space.

E5) Let $Q' = \{ax^2 + bx + c | a \neq 0, a, b, c \in C\}$
Why is Q' not a complex vector space under the usual operations?

E6) Let $V \subseteq R^3$ be given by
$V = \{(x, y, z) | ax + by + cz = 0$, for fixed $a, b, c \in R\}$.
Prove that V is a real vector space.

E7) Show that C^n is a complex vector space.

Note: We often drop the mention of the underlying field of a vector space if it is understood. For example, we may say that "$\mathbf{R}^n$ is a vector space" when we mean that "$\mathbf{R}^n$ is a vector space over $\mathbf{R}$".

Now let us look more closely at vector spaces.

3.3 FURTHER PROPERTIES OF A VECTOR SPACE

The examples and exercises in the last section illustrate different vector spaces. Elements of a vector space may be directed line segments, or ordered pairs of real numbers, or polynomials, or functions. The one thing that is common in all these examples is that each is a vector space; in each there is an addition and a scalar multiplication with the same basic properties VS1 – VS10. In this section, from these properties we develop some other useful properties which all vector spaces have.

Before we proceed, let us make a remark about notation. For a vector space V over a field F, its additive identity will be called the zero vector and, will be denoted by **0** (in bold face), to avoid confusing it with the element 0 in F, which is a scalar.

We will now state and prove some properties of a vector space.

Theorem 1: Let V be any vector space over F. Then

a) $\alpha \mathbf{0} = \mathbf{0}$, for all scalars α.

b) $0 v = \mathbf{0}$ for all $v \in V$.

c) $(-\alpha)v = -(\alpha v)$ $\alpha \in F$, $v \in V$.

Proof: a) By VS7, $\forall$ u, $v \in V$, and $\alpha \in F$.
$\alpha(u + v) = \alpha u + \alpha v$.
In particular, $\alpha(\mathbf{0} + \mathbf{0}) = \alpha \cdot \mathbf{0} + \alpha \cdot \mathbf{0}$.
But $\mathbf{0} + \mathbf{0} = \mathbf{0}$, as $\mathbf{0}$ is the additive identity of V.
Hence, $\alpha \cdot \mathbf{0} + \alpha.\mathbf{0} = \alpha \cdot \mathbf{0}$.
Adding the additive inverse $-(\alpha.\mathbf{0})$, of $\alpha.\mathbf{0}$, to both sides we get $\alpha.\mathbf{0} = \mathbf{0}$.

b) As $(\alpha + \beta) v = \alpha v + \beta v$, $\forall$ scalars α, β and $\forall v \in V$, we see that, $\forall v \in V$, $0.v = (0 + 0)v = 0.v + 0.v$. Adding $-(0.v)$ to both sides, we get $\mathbf{0} = 0.v$.

c) By VS8, for any $\alpha \in F$ and $v \in V$, $(\alpha + (-\alpha)) v = \alpha v + (-\alpha)v$. Also, by (b) above, $(\alpha + (-\alpha)) v = 0.v = \mathbf{0}$. These two equations give us $\alpha v + (-\alpha)v = \mathbf{0}$. Thus, by VS4, $(-\alpha) v = -(\alpha v)$.

Using Theorem 1 we prove the following result.

Corollary: Prove that, for $\alpha \in F$ and $u \in V$, $\alpha u = \mathbf{0}$ if and only if either $\alpha = 0$ or $u = \mathbf{0}$.

Proof: Suppose $\alpha = 0$, then $\alpha u = \mathbf{0}$ by (b) of Theorem 1. Suppose $u = \mathbf{0}$ then $\alpha u = \alpha.\mathbf{0} = \mathbf{0}$ by (a) of Theorem 1. Hence $\alpha = 0$ or $u = \mathbf{0} \Rightarrow \alpha u = \mathbf{0}$.

Conversely, suppose $\alpha.u = \mathbf{0}$. If $\alpha = 0$, there is nothing to prove. If $\alpha \neq 0$, then $1/\alpha$ exists and is a scalar, and we have
$u = 1.u = (1/\alpha.\alpha)\, u = 1/\alpha(\alpha u) = 1/\alpha(\mathbf{0}) = \mathbf{0}$.

That is, $\alpha u = \mathbf{0} \Rightarrow \alpha = 0$ or $u = \mathbf{0}$.

We give some more properties in the form of exercises for you to try.

E E8) Prove that, in a vector space V, $\forall u \in V$, $(-1)(-u) = u$.

E E9) Prove that $\forall$ u, v in a vector space, $-u-v = -(u+v)$

E10) Prove that $-(-u) = u$, $\forall$ u in a vector space.

E11) Prove that $\alpha(u-v) = \alpha u - \alpha v$ for all scalars α and $\forall$ u,v in a vector space.

Let us now look at some subsets of the underlying sets of vector spaces

3.4 SUBSPACES

In E3 you saw that V, a subset of C, was also a vector space. You also saw, in Example 6, that the subset

$V = \{(x,y) \in R^2 | y = 5x\}$,

of the vector space R^2, is itself a vector space under the same operations as those in R^2. In these cases V is a subspace of R^2. Let us see what this means.

Definition: Let V be a vector space and $W \subseteq V$. If W is also a vector space under the same operations as those in V, we say that W is a **subspace** of V.

The following theorem gives the criterion for a subset to be a subspace.

Theorem 2: A non-empty subset W, of a vector space V over a field F, is a subspace of V provided

a) $w_1 + w_2 \in W$, $\forall w_1, w_2 \in W$

b) $\alpha w \in W \; \forall \alpha \in F$ and $w \in W$.

c) **0**, the additive identity of V, also belongs to W.

Proof: We have to show that the properties VS1 – VS10 hold for W.

VS1 is true because of (a) given above.

VS2 and VS5 are true for elements of W because they are true for elements of V.

VS3 is true because of (c) above.

VS4 is true because, if $w \in W$ then $(-1)w = -w \in W$, by (b) above.

VS6 is true because of (b) above.

VS7 to VS10 hold true because they are true for V.

Therefore, W is a vector space in its own right, and hence, it is a subspace of V.

The next theorem says that condition (c) in Theorem 2 is unnecessary.

Theorem 3: A non-empty subset W, of a vector space V over a field F, is a subspace of V if and only if

a) $w_1 \in W, w_2 \in W \Rightarrow w_1 + w_2 \in W$

b) $\alpha \in F, w \in W \Rightarrow \alpha w \in W$.

A non-empty subset of a vector space is a subspace iff it is closed under vector addition and scalar multiplication.

Proof: If W is a subspace, then obviously (a) and (b) are satisfied.

Conversely, suppose (a) and (b) are satisfied. To show that W is a subspace of V, Theorem 2 says that we only need to prove that $0 \in W$. Since W is non-empty, there is some $w \in W$. Then, by (b), $0 \cdot w \in W$, i.e., $0 \in W$.

This completes the proof of the theorem.

Actually both the conditions in Theorem 3 can be merged to give the following compact result.

Theorem 4: A non-empty subset W, of a vector space V over the field **F**, is a subspace of V if and only if
$\alpha w_1 + \beta w_2 \in W \ \forall\ \alpha, \beta \in \mathbf{F}$ and $w_1, w_2 \in W$.

Proof: Firstly, suppose W is a subspace of V. Then, by Theorem 3, for any $\alpha, \beta \in \mathbf{F}$ and $w_1, w_2 \in W$, we have $\alpha w_1 \in W$ and $\beta w_2 \in W$, so that $\alpha w_1 + \beta w_2 \in W$.

Conversely, suppose $\alpha w_1 + \beta w_2 \in W \ \forall\ \alpha, \beta \in \mathbf{F}$ and $w_1, w_2 \in W$. Then, in particular, for $\alpha = 1 = \beta$ (remember $1 \in \mathbf{F}$), $w_1 + w_2 \in W$. Also, if we put $\beta = 0$ in $\alpha w_1 + \beta w_2$, we get $\alpha w_1 \in \Xi \ \forall\ \alpha \in \mathbf{F}$ and $w_1 \in W$. $\therefore$, by Theorem 3, W is a subspace.

Hence, the theorem is proved.

Let us use this theorem to obtain some more examples of vector spaces.

Example 7: Prove that the subset
$W = \{(x, 2x, 3x) | x \in \mathbf{R}\}$
of $\mathbf{R}^3$ is a subspace of $\mathbf{R}^3$.

Solution: If we take $x = 0$, we see that $(0,0,0) \in W$, so $W \neq \emptyset$. (Remember $\emptyset$ denotes the empty set.)

Next, $w_1 \in W, w_2 \in W \Rightarrow w_1 = (x,2x,3x), w_2 = (y,2y,3y)$, where $x \in \mathbf{R}, y \in \mathbf{R}$. Thus
$\alpha w_1 = (\alpha x, 2\alpha x, 3\alpha x)$ and $\beta w_2 = (\beta y, 2\beta y, 3\beta y)$, for $\alpha, \beta \in \mathbf{R}$.
$\Rightarrow \alpha w_1 + \beta w_2 = (\alpha x + \beta y, 2(\alpha x + \beta y), 3(\alpha x + \beta y))$
$\Rightarrow \alpha w_1 + \beta w_2 = (z, 2z, 3z)$, where $z = \alpha x + \beta y \in \mathbf{R}$.
$\Rightarrow \alpha w_1 + \beta w_2 \in W$.

Hence, by Theorem 4, W is a subspace of $\mathbf{R}^3$.

Example 8: Which of the following subsets W of $\mathbf{R}^4$ are subspaces of $\mathbf{R}^4$?
The set of all $w = (x_1,x_2,x_3,x_4) \in \mathbf{R}^4$ such that

a) $x_1 = 0$, (b) $x_2 = 1$, (c) $x_3 < 0$, (d) $2x_1 + 5x_4 = 0$.

Solution: a) Here, $W = \{(0,x_2,x_3,x_4) | x_2,x_3,x_4 \in \mathbf{R}\}$
Obviously, $W \neq \phi$ as $(0,0,0,0) \in W$.
Next, $w_1, w_2 \in W \Rightarrow \quad w_1 = (0,x_2,x_3,x_4), x_i \in \mathbf{R}$, for $i = 2,3,4$.
$\qquad\qquad w_2 = (0,y_2,y_3,y_4), y_i \in \mathbf{R}$, for $i = 2,3,4$.
$\Rightarrow \alpha w_1 = (0, \alpha x_2, \alpha x_3, \alpha x_4)$ and $\beta w_2 = (0, \beta y_2, \beta y_3, \beta y_4)$, $\alpha,\beta \in \mathbf{R}$.
$\Rightarrow \alpha w_1 + \beta w_2 = (0, \alpha x_2 + \beta y_2, \alpha x_3 + \beta y_3, \alpha x_4 + \beta y_4) \in W$.
Hence W is a subspace of $\mathbf{R}^4$.

(b) Here, $W = \{(x_1,1,x_3,x_4) | x_1,x_3,x_4 \in \mathbf{R}\}$

Again $W \neq \phi$, as $(1,1,1,1) \in W$.

Now $w_1 \in W, w_2 \in W \Rightarrow w_1 = (x_1,1,x_3,x_4), w_2 = (y_1,1,y_3,y_4)$
$\Rightarrow w_1 + w_2 = (x_1 + y_1, 2, x_3 + y_3, x_4 + y_4)$
$\Rightarrow w_1 + w_2 \notin W$
So W is not a subspace of $\mathbf{R}^4$

Note: An easier proof for (b) would be:

$0 \cdot w = (0,0,0,0) \notin W$; $\therefore$, W is not a subspace.

c) Here, $W = \{(x_1,x_2,x_3,x_4) | x_i \in \mathbf{R}, x_3 < 0\}$
Then $W \neq \phi$, as $(0,0,-1,0) \in W$.

Now, $w = (0,0,-1,0) \in W$, but $(-1)\,w = -w = (0,0,1,0) \notin W$.
Therefore, W is not a subspace of $\mathbb{R}^4$.

d) Now, $W = \{(x_1, x_2, x_3, x_4) | x_i \in \mathbb{R}, 2x_1 + 5x_4 = 0\}$
Obviously $(0,0,0,0) \in W$, so $W \neq \phi$. Next,
$w_1 \in W, w_2 \in W \Rightarrow w_1 = (x_1, x_2, x_3, x_4)$ with $2x_1 + 5x_4 = 0$
and $w_2 = (y_1, y_2, y_3, y_4)$ with $2y_1 + 5y_4 = 0$
$\Rightarrow w_1 + w_2 = (x_1+y_1, x_2+y_2, x_3+y_3, x_4+y_4)$ with $2(x_1+y_1)+5(x_4+y_4)$
$= (2x_1 + 5x_4) + (2y_1 + 5y_4) = 0 + 0 = 0.$
$\Rightarrow w_1 + w_2 \in W.$

Finally,
$\alpha \in \mathbf{R}, w \in W \Rightarrow \alpha \in \mathbf{R}, w = (x_1, x_2, x_3, x_4)$ with $2x_1 + 5x_4 = 0$.
$\Rightarrow \alpha\, w = (\alpha x_1, \alpha x_2, \alpha x_3, \alpha x_4)$ with $2(\alpha x_1) + 5(\alpha x_4) = \alpha(2x_1 + 5x_4) = 0.$
$\Rightarrow \alpha\, w \in W.$

So W is a subspace of $\mathbf{R}^4$.

Note: We could have also solved (d) by using Theorem 4 as follows:
For $\alpha, \beta \in \mathbf{R}$ and (x_1,x_2,x_3,x_4), (y_1,y_2,y_3,y_4) in W we have
$\alpha(x_1,x_2,x_3,x_4) + \beta(y_1,y_2,y_3,y_4) = (\alpha x_1+\beta y_1, \alpha x_2+\beta y_2, \alpha x_3+\beta y_3, \alpha x_4+\beta y_4)$
with $2(\alpha x_1+\beta y_1) + 5(\alpha x_4+\beta y_4) = \alpha(2x_1 + 5x_4) + \beta(2y_1+5y_4) = 0.$

Thus, $\alpha, \beta \in \mathbf{R}$ and $w_1, w_2 \in W \Rightarrow \alpha w_1 + \beta w_2 \in W$.
This shows that W is a subspace of $\mathbf{R}^4$.

Example 9: Let V be a vector space over F and $v \in V$.
Show that the subset $F\,v = \{\alpha v | \alpha \in F\}$ is a subspace of V.

All scalar multiples of a fixed vector form a subspace.

Solution: $F\,v \neq \phi$ because $0 \cdot v = 0 \in F\,v$.
Now, if $\alpha\, v$ and $\beta\, v \in F\,v$ then $\alpha v + \beta v = (\alpha + \beta)\, v \in Fv$.
Also, $\alpha \in F$, and $\beta\, v \in F\,v \Rightarrow \alpha\,(\beta v) = (\alpha\beta)\, v \in Fv$, since $\alpha\beta \in F$.

Thus, by Theorem 3, F v is a subspace of V.

Note: The subspace $\mathbf{R}v$, of $\mathbf{R}^n$, represents a line in n-dimensional space.

E12) Prove that $W = \{(x_1,x_2,x_3) \in \mathbf{R}^3 | 2x_1 + x_2 - x_3 = 0\}$ is a subspace of $\mathbf{R}^3$.

E13) For each of the following subsets W of $\mathbb{R}^3$, determine whether it is a subspace of $\mathbf{R}^3$.
W is the set of those vectors (x_1, x_2, x_3) in $\mathbb{R}^3$ such that
(a) $x_1 = -x_2$; (b) $x_1^2 \geq 0$; (c) $x_1 x_2 = 0$; (d) $x_1 + x_2 + x_3 = 1$.

E E14) Show that {0} is a subspace of the vector space V over F.

In Example 9 you saw that an element $v \in V$ gives rise to a subspace of V. In the next section we look at such subspaces of V, which grow out of subsets of V that are much smaller than the concerned subspace.

3.5 LINEAR COMBINATIONS

In Unit 2 you came across the fact that any element of $\mathbb{R}^3$ could be written as $a \cdot i + b \cdot j + c \cdot k$, where $a,b,c \in \mathbb{R}$. In this section we will generalise this. Consider the following definition.

Definition: If $v_1, v_2, \ldots, v_n$ are elements of a vector space over F, and $\alpha_1, \alpha_2, \ldots \alpha_n \in F$, then the vector

$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$

is called a **linear combination** of the vectors $v_1, v_2, \ldots, v_n$, or of the set $\{v_1, v_2, \ldots, v_n\}$.

For instance, since

$(2,4,3) = 2(1,-1,0) + 3(0,2,1)$, $(2,4,3)$ is a linear combination of $(1,-1,0)$ and $(0,2,1)$

We are now ready to generalise the result of Example 9.

Theorem 5: If $v_1, v_2, \ldots, v_n$ belong to a vector space V over a field F, then
$W = \{\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n | \alpha_i \text{ are scalars}\}$
is a subspace of V.

Proof: Firstly, 0 is a scalar and $0 \cdot v_1 + 0 \cdot v_2 + \ldots + 0 \cdot v_n$
$= 0 + 0 + \ldots + 0$
$= 0.$
So $0 \in W$, and $W \neq \phi$
Secondly, $w_1 \in W$, $w_2 \in W$

$\Rightarrow w_1 = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = \sum_{i=1}^{n} \alpha_i v_i, \alpha_i \in F$

and $w_2 = \beta_1 v_1 + \beta_2 v_2 + \beta_n v_n = \sum_{i=1}^{n} \beta_i v_i \in F.$

$\Rightarrow w_1 + w_2 = (\alpha_1 + \beta_1) v_1 + \ldots + (\alpha_n + \beta_n) v_n \Rightarrow w_1 + w_2 \in W$
Finally, if α is a scalar, and $w \in W$, we have
$w = \alpha_1 v_1 + \ldots + \alpha_n v_n$, where α_i is a scalar $\forall i = 1, \ldots, n$.
$\Rightarrow \alpha w = (\alpha\alpha_1) v_1 + (\alpha\alpha_2) v_2 + \ldots + (\alpha\alpha_n) v_n$.
$\Rightarrow \alpha w \in W$
This proves the theorem.

We often denote W (in Theorem 5) by $F v_1 + \ldots + F v_n$.

Let us look at the vector space $\mathbf{R}^n$, over $\mathbf{R}$. In this, we see that every vector is a linear combination of the n vectors $e_1 = (1,0,\ldots,0)$, $e_2 = (0,1,0,\ldots,0), \ldots, e_n = (0,\ldots,0,1)$. This is because $(x_1,\ldots,x_n) = x_1e_1 + x_2e_2 + \ldots + x_ne_n$, $x_i \in \mathbf{R}$. In this case we say that the set $\{e_1,\ldots,e_n\}$ spans $\mathbf{R}^n$. Let us see what spanning means.

Definition: Let V be a vector space over F, and let $S \subseteq V$. The **linear span** of S is defined to be the set of all linear combinations of a finite number of elements of S. It is denoted by [S]. Thus,

$[S] = \{ \sum_{i=1}^{n} \alpha_i v_i \mid n \text{ positive integer}, v_i \in S, \alpha_i \text{ scalars}\}$

We also say that S **generates** [S].
Note that S is only a subset of V, and not necessarily a subspace of V. Also note that [S] is the set of finite sums of the form $\alpha_1 v_1 + \ldots + \alpha_n v_n$, where $\alpha_i \in F$ and $v_i \in S$.

Example 10: Suppose $S \subseteq \mathbf{R}^2$, $S = \{(1,0), (0,1)\}$. What is [S]?

Solution: $[S] = \{\alpha(1,0) + \beta(0,1) \mid \alpha,\beta \in \mathbf{R}\}$, i.e.,
$[S] = \{(\alpha,\beta) \mid \alpha,\beta \in \mathbf{R}\}$.
In this case, the linear span of S is the whole of $\mathbf{R}^2$. Thus, $\{(1,0), (0,1)\}$ generates $\mathbf{R}^2$.

Example 11: Suppose $S \subseteq \mathbf{R}^3$, $S = \{(1,-1,0)\}$. What is [S]?

Solution: $[S] = \{\alpha(1,-1,0) \mid \alpha \in \mathbf{R}\}$.
$= \{(\alpha, -\alpha, 0) \mid \alpha \in \mathbf{R}\}$

Example 12: Let P be the vector space of real polynomials, and
$S = \{x, x^2+1, x^3-1\} \subseteq P$. What is [S]?

Solution: $[S] = \{\alpha x + \beta(x^2+1) + \tau(x^3-1) \mid \alpha, \beta, \tau \in \mathbf{R}\}$.
$= \{\tau x^3 + \beta x^2 + \alpha x + (\beta-\tau) \mid \alpha, \beta, \tau \in \mathbf{R}\}$

E15) Let $S = \{1, x, x^2\}$ be a subset of P in the example above. Does $2x + 3x^3 \in [S]$?

In the examples given above you may have noticed that [S] is a subspace of V. We prove this fact now.

Theorem 6: If S is a non-empty subset of a vector space V over **F**, then [S] is a subspace of V.

Proof: Since $S \neq \phi$ and $S \subseteq [S]$, $[S] \neq \phi$. Also, since $S \subseteq V$, $[S] \subseteq V$.

Now, $s_1 \in [S]$, $s_2 \in [S]$

$\Rightarrow s_1 = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$, for $v_i \in S$, $\alpha_i \in F$ and

$s_2 = \beta_1 w_1 + \beta_2 w_2 + \ldots + \beta_m w_m$, for $w_i \in S$, $\beta_i \in F$.

Thus, for $\alpha, \beta \in F$, $\alpha s_1 = \alpha\alpha_1 v_1 + \alpha\alpha_2 v_2 + \ldots + \alpha\alpha_n v_n$,

$\beta s_2 = \beta\beta_1 w_1 + \beta\beta_2 w_2 + \ldots + \beta\beta_m w_m$

$\Rightarrow \alpha s_1 + \beta s_2 = \alpha\alpha_1 v_1 + \ldots + \alpha\alpha_n v_n + \beta\beta_1 w_1 + \ldots + \beta\beta_m w_m$, with $v_i, w_i \in S$.

and $\alpha\alpha_i \in F$, $\beta\beta_i \in F$. This shows that $\alpha s_1 + \beta s_2$ is a linear combination of a finite number of elements of S. Thus, $\alpha s_1 + \beta s_2 \in [S]$. Therefore, by Theorem 4, [S] is a subspace of V

Theorem 6 shows that the linear span of S is a subspace containing S. In fact, it is the smallest subspace of V containing S, as you will see now.

The linear span of S is the smallest subspace of V containing S.

Theorem 7: If S is a subset and T a subspace of the vector space V over **F**, such that $S \subseteq T$, then $[S] \subseteq T$.

Proof: Let $s \in [S]$, then

$$s = \sum_{i=1}^{n} a_i v_i, \text{ where } v_i \in S,\ a_i \in F$$

As $S \subseteq T$, $v_i \in T\ \forall i = 1, \ldots, n$. As T is a subspace and $v_i \in T$ for all i, $\sum_{i=1}^{n} a_i v_i \in T$, i.e., $s \in T$.

We have proved that $s \in [S] \Rightarrow s \in T$.

Hence, $[S] \subseteq T$.

An immediate corollary to Theorem 7 follows.

Corollary 1: If S is a subspace of V, then $[S] = S$.

Proof: Since S is a subspace containing S, Theorem 7 gives us $[S] \subseteq S$. But $S \subseteq [S]$ always. Therefore, $[S] = S$.

The theorems above say that we can form subspaces from mere subsets of a space. Given a subset S of a vector space V, if S is not a subspace of V, what is the 'minimum' that we must add to S to make it a subspace? The answer is – all the finite linear combinations of vectors of S.

Look at the following examples.

Example 13: Let $S = \{(1,1,0), (2,1,3)\} \subseteq \mathbf{R}^3$. Determine whether the following vectors of $\mathbf{R}^3$ are in [S].

(a) (0,0,0); (b) (1,2,3); (c) (4/3, 1,1).

Solution: $[S] = \{\alpha(1,1,0) + \beta(2,1,3) \mid \alpha, \beta \in \mathbf{R}\}$

$= \{(\alpha + 2\beta, \alpha + \beta, 3\beta) \mid \alpha, \beta \in \mathbf{R}\}$

(a) $(0,0,0) \in [S]$, since [S] is a subspace and (0,0,0) is the additive identity of $\mathbf{R}^3$.

(b) $(1,2,3) \in [S]$ if we can find $\alpha, \beta \in \mathbf{R}$, such that $(\alpha + 2\beta, \alpha + \beta, 3\beta) = (1,2,3)$, i.e., $\alpha + 2\beta = 1$, $\alpha + \beta = 2$, $3\beta = 3$,

Now $3\beta = 3 \Rightarrow \beta = 1$, and then, $\alpha + \beta = 2 \Rightarrow \alpha = 1$.

But then $\alpha + 2\beta = 1 + 2 = 3 \neq 1$. Hence, $(1,2,3) \notin [S]$.

(c) $(4/3,1,1) \in [S]$ if $\alpha + 2\beta = 4/3$, $\alpha + \beta = 1$, $3\beta = 1$ for some $\alpha, \beta \in \mathbf{R}$.

These equations are satisfied if $\beta = 1/3$, $\alpha = 2/3$.

So $(4/3,1,1) \in [S]$.

E16) If $S = \{(1,2,1), (2,1,0)\} \subseteq \mathbf{R}^3$, determine whether the following vectors of $\mathbf{R}^3$ are in [S].
(a) (5,5,1), (b) = (2,1,0). (c) (4,5,2).

E17) Let **P** be the vector space of polynomials over **R** and $S = \{x, x^2+1, x^3-1\}$. Determine whether the following polynomials are in [S].
(a) $x^2 + x + 1$, (b) $2x^3+x^2+3x+2$.

Now that you have got used to the concept of subspaces we go on to construct new vector spaces from existing ones.

3.6 ALGEBRA OF SUBSPACES

In this section we will consider the union, intersection, sum and direct sum of vector spaces.

3.6.1 Intersection

If U and W are subspaces of a vector space V over a field **F**, then the set $U \cap W$ is a subset of V. We will prove that it is actually a subspace of V.

Theorem 8: The intersection of two subspaces is a subspace.

Proof: Let U and W be two subspaces of a vector space V. Then $\mathbf{0} \in U$ and $\mathbf{0} \in W$. Therefore, $\mathbf{0} \in U \cap W$; hence $U \cap W \neq \phi$.

Next, if $v_1 \in U \cap W$, and $v_2 \in U \cap W$, then $v_1 \in U, v_2 \in U, v_1 \in W, v_2 \in W$.

Thus, for any $\alpha,\beta \in F$, $\alpha v_1 + \beta v_2 \in U$, $\alpha v_1 + \beta v_2 \in W$ (as U and W are subspaces).

$\therefore$, $\alpha v_1 + \beta v_2 \in U \cap W$.

This proves that $U \cap W$ is a subspace of V.

Example 14: $U = \{(x,2x,3x) | x \in \mathbb{R}\}$ and
$W = \{(0,y,(3/2)y) | y \in \mathbb{R}\}$ are subspaces of $\mathbb{R}^3$. What is $U \cap W$?

Solution: Any element of $U \cap W$ is of the form $(x,2x,3x)$ and of the form $(0,y,(3/2)y)$. Thus, the only possibility is $(0,0,0)$. Therefore, $U \cap W = \{(0,0,0)\}$. By E 14 you know that this is a vector space.

Example 15: $U = \{(x,y,0) | x,y \in \mathbb{R}\}$ and
$W = \{(0,y,z) | y,z \in \mathbb{R}\}$ are subspaces of $\mathbb{R}^3$. What is $U \cap W$?

Solution: $U \cap W$ is the set $\{(0,y,0) | y \in \mathbb{R}\}$.
In this example note that U is the xy-plane, W is the yz-plane and $U \cap W$ is the y-axis.

E E18) If $U = \{(x,y,2x) | x,y \in \mathbb{R}\}$ and $W = \{(x,2x,y) | x,y \in \mathbb{R}\}$, what is $U \cap W$?

Note: It can be shown that the intersection of any finite or infinite family of subspaces is a subspace. In particular, if $V_1, V_2, \ldots, V_n$ are all subspaces of V, then $V_1 \cap V_2 \cap \ldots \cap V_n$ is a subspace of V.

Let us now look at what happens to the union of two or more subspaces.

3.6.2 Sum

Consider the subspaces U and W of $\mathbb{R}^3$ given in Example 15. Here $v_1 = (1,2,0) \in U$ and $v_2 = (0,2,3) \in W$. Therefore, v_1 and v_2 belong to $U \cup W$. But $v_1 + v_2 = (1,4,3)$ is neither in U nor in W, and hence, not in $U \cup W$. So $U \cup W$ is not a subspace of $\mathbb{R}^3$. Thus, we see that, while the intersection of two subspaces is a subspace, the union of two subspaces may not be a subspace. However, if we take two subspaces U and W, of a vector space V, then $[U \cup W]$, the linear span of $U \cup W$, is a subspace of V.

What are the elements of $[U \cup W]$? They are linear combinations of elements of $U \cup W$. So, for each $v \in [U \cup W]$, there are vectors $v_1, v_2, \ldots, v_n \in U \cup W$ of which v is a linear combination. Now some (or all) of the $v_1, \ldots, v_n$ are in U and the rest in W. We rename those that are in U as $u_1, u_2, \ldots, u_j$ and those in W as $w_1, w_2, \ldots, w_k$ $(j \geq 0, k \geq 0, j + k = n)$.

Then, there are scalars $\alpha_1, \ldots, \alpha_j, \beta_1, \ldots, \beta_k$ such that

$$v = \alpha_1 u_1 + \ldots + \alpha_j u_j + \beta_1 w_1 + \ldots + \beta_k w_k$$
$$= u + w,$$

where $u = \alpha_1 u_1 + \ldots + \alpha_j u_j \in U$, since each $u_i \in U$, and $w = \beta_1 w_1 + \ldots + \beta_k w_k \in W$, since each $w_i \in W$. (If $j = 0$, we take $u = \mathbf{0}$; if $k = 0$, we take $w = \mathbf{0}$.) So what we have proved is that every element of $[U \cup W]$ is of the type $u+w$, $u \in U$, $w \in W$. This motivates the following definition.

Definition: If A and B are subsets of a vector space, we define the set $A + B$ by $A + B = \{a+b | a \in A, b \in B\}$.

Thus, each element of $A + B$ is the sum of an element of A and an element of B.

Example 16: If $A = \{(0,0), (1,1), (2,-3)\}$ and $B = \{(-3,1)\}$ are subsets of $\mathbb{R}^2$, find $A + B$.

Solution: $A+B = \{(-3,1), (-2,2), (-1,-2)\}$ because, for example,
$(0,0) + (-3,1) = (-3,1)$
$(1,1) + (-3,1) = (-2,2)$, etc.

Example 17: Let $A = \{(0,y,z) | y,z \in \mathbf{R}\}$ and $B = \{(x,0,z) | x,z \in \mathbf{R}\}$.
Prove that $A+B = \mathbf{R}^3$.

Solution: Since $A \subseteq \mathbf{R}^3$, $B \subseteq \mathbf{R}^3$, so $A+B \subseteq \mathbf{R}^3$. It is, therefore, enough to prove that $\mathbf{R}^3 \subseteq A + B$. Let $(a,b,c) \in \mathbf{R}^3$. Then
$(a,b,c) = (0,b,c/2) + (a,0,c/2)$, where $(0,b,c/2) \in A$ and $(a,0,c/2) \in B$.
So $(a,b,c) \in A+B$.
Thus, $\mathbf{R}^3 \subseteq A+B$.
Hence, $A + B = \mathbf{R}^3$.

Note that in the discussion preceding the definition of a sum of subsets, we have actually proved that if U and W are subspaces of a vector space V, then $[U \cup W] \subseteq U + W$. Indeed, we have the following theorem.

Theorem 9: If A and B are subspaces of a vector space V, then $[A \cup B] = A + B$.

Proof: We have already proved (see above) that $[A \cup B] \subseteq A + B$. So it only remains to prove that $A + B \subseteq [A \cup B]$.

Let $v \in A + B$, then $v = a+b$, $a \in A$, $b \in B$. Now $a \in A \Rightarrow a \in A \cup B \Rightarrow a \in [A \cup B]$.

Similarly, $b \in B \Rightarrow b \in A \cup B \Rightarrow b \in [A \cup B]$. As $[A \cup B]$ is a vector space and $a, b \in [A \cup B]$, we see that $a+b \in [A \cup B]$, i.e., $v \in [A \cup B]$. This completes the proof of the theorem.

Since $[A \cup B]$ is the smallest subspace containing $A \cup B$, we see, from Theorem 9 that A+B is the smallest subspace of V containing both A and B.

E19) For the subspaces $A = \{(x,0,0) | x \in \mathbf{R}\}$ and $B = \{(0,y,0) | y \in \mathbf{R}\}$ of $\mathbf{R}^3$, find $[A \cup B]$.

We consider a special kind of sum of subsets now.

3.6.3 Direct Sum

If A and B are subspaces of a vector space, you know that every vector v in A+B is of the form a+b, where $a \in A$, $b \in B$. But in how many ways can a given $v \in A+B$ be expressed in the form, a + b?

In Example 17 we have expressed $(a,b,c) \in \mathbf{R}^3$ in the form a+b by writing
$(a,b,c) = (0,b,c/2) + (a,0,c/2)$.

But we could also write

$(a,b,c) = (0,b,0) + (a,0,c)$
or $(a,b,c) = (0,b,c) + (a,0,0)$
or $(a,b,c) = (0,b,c/3) + (a,0,2c/3)$.

Indeed, for any real number δ we can write $(a,b,c) = (0,b,\delta) + (a,0,c-\delta)$. Note that, in each case, we have expressed (a,b,c) as a sum of vector from A and a vector from B. So, in this case, there are infinitely many ways of writing $v \in A + B$ in the form a+b, with $a \in A$, $b \in B$

But there are some cases in which every vector $v \in A+B$ can be written in one and only one way as $a+b$, $a \in A$, $b \in B$. For example, suppose $A = \{(x,y,0)|x,y \in \mathbf{R}\}$ and $B = \{(0,0,z)|z \in \mathbf{R}\}$.

Then, for any $(p,q,r) \in \mathbf{R}^3$ we can write

$(p,q,r) = (p,q,0) + (0,0,r) \in A + B$.

It follows that $A + B = \mathbf{R}^3$. But here (p,q,r) can be written in only one way as $a+b$, namely $(p,q,0) + (0,0,r)$, because, if we write $(p,q,r) = (x,y,0) + (0,0,z)$, then $(p,q,r) = (x,y,z)$, so that $x = p$, $y = q$, $z = r$. This means that $(x,y,0) = (p,q,0)$ and $(0,0,z) = (0,0,r)$.

Now, note that in this case $A \cap B = \{(0,0,0)\}$, whereas in the earlier example
$A \cap B = \{(0,0,z)|z \in \mathbf{R}\} \neq \{(0,0,0)\}$
It is this difference in $A \cap B$ that is reflected in a unique or a multiple representation o v in the form $a+b$.

Definition: Let A and B be subspaces of a vector space. The sum $A + B$ is said to be the **direct sum** of A and B (and is denoted by $A \oplus B$) if $A \cap B = \{0\}$.

We have the following result.

Theorem 10: A sum $A + B$, of subspaces A and B, is a direct sum $A \oplus B$ if and only if every $v \in A+B$ is uniquely expressible in the form $a + b$, $a \in A$, $b \in B$

Proof: First suppose $A + B$ is a direct sum i.e., $A \cap B = \{0\}$. If possible, suppose v has two representations,

$v = a_1 + b_1$ and $v = a_2 + b_2$, $a_i \in A$, $b_i \in B$.
Then $a_1 + b_1 = a_2 + b_2$, i.e., $a_1 - a_2 = b_2 - b_1$.
Now $a_1, a_2 \in A \Rightarrow a_1 - a_2 \in A$. Similarly, $b_2 - b_1 \in B$, that is,
$a_1 - a_2 \in B$ (since $a_1 - a_2 = b_2 - b_1$).
Thus, $a_1 - a_2 \in A \cap B \Rightarrow a_1 - a_2 = 0 \Rightarrow a_1 = a_2$.
And then, $b_1 = b_2$.
This means that $a_1 + b_1$ and $a_2 + b_2$ are the same representations of v as $a + b$.

Conversely, suppose every $v \in A+B$ has exactly one representation as $a+b$. We must prove that $A \cap B = \{0\}$.

Since A and B are subspaces, $0 \in A$, $0 \in B$. $\therefore \{0\} \in A \cap B$.

If $A \cap B \neq \{0\}$, then there must be some $v \neq 0$ such that $v \in A \cap B$.

Then, v has two distinct representations as $a+b$,
namely, $v + 0$ ($v \in A$, $0 \in B$) and $0+v$ ($0 \in A$, $v \in B$). This is a contradiction. So $A \cap B = \{0\}$. Hence $A+B$ is a direct sum.

Example 18: Let A and B be subspaces of $\mathbf{R}^3$ defined by
$A = \{(x,y,z) \in \mathbf{R}^3 | x = y = z\}$, $B = \{(0,y,z)|y,z \in \mathbf{R}\}$.
Prove that $\mathbf{R}^3 = A \oplus B$.

Solution: First note that $A + B \subseteq \mathbf{R}^3$. Secondly, if $(a,b,c) \in A \cap B$, then $a = b = c$, anc $a = 0$; so $a = 0 = b = c$, i.e., $(a,b,c) = (0,0,0)$. Hence, the sum $A + B$ is the direct sum $A \oplus B$. Next given any $(a,b,c) \in \mathbf{R}^3$, we have $(a,b,c) = (a,a,a) + (0,b-a,c-a)$, where $(a,a,a) \in A$ and $(0,b-a,c-a) \in B$; this proves that $\mathbf{R}^3 \subseteq A \oplus B$. Therefore, $\mathbf{R}^3 = A \oplus B$.

A function $f: \mathbf{R} \to \mathbf{R}$ is called an **even function** if $f(x) = f(-x)$ **$\forall x \in \mathbf{R}$, and an odd function if** $f(-x) = -f(x) \forall x \in \mathbf{R}$.

Example 19: Let V be the space of all functions from $\mathbf{R}$ to $\mathbf{R}$, and A and B be the subspaces of V defined by
$A = \{f | f(x) = f(-x), \forall x\}$
$B = \{f | f(-x) = -f(x), \forall x\}$
i.e., A is the subspace of all even functions and B is the subspace of all odd functions. Show that $V = A \oplus B$.

Solution: First, suppose $f \in A \cap B$, then $\forall x \in R$, $f(-x) = f(x)$ and $f(-x) = -f(x)$. So, $\forall x$, $f(x) = -f(x)$, i.e., $\forall x$, $f(x) = 0$. Thus, f is the zero function, and $A \cap B = \{0\}$.

Next, let $f \in V$, define

$g(x) = \frac{1}{2}\{f(x) + f(-x)\}$, and

$h(x) = \frac{1}{2}\{f(x) - f(-x)\}$.

Then, (i) $f(x) = g(x) + h(x) \ \forall x \in R$, i.e., $f = g + h$.

(ii) $g(-x) = \frac{1}{2}\{f(-x) + f(x)\} = g(x)$. ∴ $g \in A$.

(iii) $h(-x) = \frac{1}{2}\{f(-x) - f(x)\} = -h(x)$. ∴ $h \in B$.

Thus, for each $f \in V$, $f = g+h$, for some $g \in A$, $h \in B$.

$\Rightarrow$ $V = A+B$, and, as $A \cap B = \{0\}$, we get

$V = A \oplus B$

Note: Example 19 says that every function from **R** to **R** can be uniquely written as the sum of an even function and an odd function.

E20) Let A,B,C be the subspaces of R^3 given by

$A = \{(x,y,z) \in R^3 | x+y+z = 0\}$,

$B = \{(x,y,z) \in R^3 | x = z\}$, $C = \{(0,0,z) | z \in R\}$

Prove that $R^3 = A+C$ and $R^3 = B+C$.

Which of these sums is/are direct?

E21) Consider the real vector space C, of all complex numbers (Example 5). If A and B are the subspaces of C given by $A = \{a+i.0 | a \in R\}$, $B = \{ib | b \in R\}$, prove that $C = A \oplus B$.

Now, we will look at vector spaces that are obtained by "taking the quotient" of a vector space by a subspace.

3.7 QUOTIENT SPACES

From a vector space V, and its subspace W, we will now create a new vector space. For this, we first define the concept of a coset.

3.7.1 Cosets

Let W be a subspace of V. If $v \in V$, the set $v + W$, defined by
$v + W = \{v + w : w \in W\}$
is called a coset of W in V.

Example 20: Consider the subspace $W = \{(a,0,0) | a \in R\}$ of R^3.
Let $v = (1,0,2)$. Find the coset $v + W$. Is it a subspace of R^3?

Solution: $v + W = \{v+w | w \in W\}$
$= \{(1,0,2) + (a,0,0) | a \in R\}$
$= \{(a+1, 0,2) | a \in R\}$
Thus, $v + W = \{(a,0,2) | a \in R\}$.
(because, as a takes all the real values, a + 1 also takes all the real values, so that we may replace a+1 by a).
v+W is not a subspace of R^3 as $(0,0,0) \notin v+W$.

Observe that each element v of V yields a coset v + W of W. Every coset of W in V is a subset of V, but it may not be a subspace of V, as you have seen in Example 20.

Example 21: With W as in Example 20, and $v = (2,0,0)$, prove that v+W is a subspace and, in fact, v+W = W.

Solution: Here $v + W = \{(2,0,0) + (a,0,0) | a \in R\}$
$= \{(a + 2, 0, 0) | a \in R\}$
$= \{(p,0,0) | p \in R\}$
$= W$

Observe that, in the example above $v \in W$ whereas in the previous example, $v \notin W$. In the next theorem we substantiate this observation.

Theorem 11: Let W be a subspace of a vector space V. Then $v \in W$ if and only if $v + W = W$. Also, if $v \notin W$, then v + W is not a subspace of V.

Proof: We first prove that $v \in W \Rightarrow v+W = W$. For this let $u \in v + W$. Then, for some $w \in W$, $u = v + w$. This implies that $u \in W$, as both $v,w \in W$. This means that $v + W \subseteq W$.
Also, $w \in W \Rightarrow w-v \in W$, since $v \in W$.
$\Rightarrow w = v+(w-v) \in v + W$, so that $W \subseteq v+W$.
This proves that v + W = W.

Now let us prove the converse, namely, $v+W = W \Rightarrow v \in W$. For this we use the fact that $0 \in W$. Then we have $v = v+0 \in v+W = W \Rightarrow v \in W$.

Lastly, we prove that, if $v \notin W$ then v+W is not a subspace of V. If v + W is a subspace of V, then $0 \in v + W$. Therefore, for some $w \in W$, $v + w = 0$, i.e., $w = -v$. Hence, $-v \in W$ and, as W is a subspace, $v \in W$.
Thus, v + W is a subspace of $V \Rightarrow v \in W$. So, $v \notin W \Rightarrow v + W$ is not a subspace of V.

v+W is a subspace of $V \Rightarrow v \in W$.

E E22) Let $W = \{f(x) \in P | f(1) = 0\}$ be a subspace of P, the real vector space of all polynomials in one variable x.
a) If $v = (x-1)(x^2+1)$, what is v + W?
b) If $v = (x-2)(x^2+1)$, what is v + W?

Now we ask : Given a vector space V and a subspace W, can we get V back if we know all the cosets of W? The answer is given in the following theorem.

Theorem 12: If W is a subspace of V, the union of all the cosets of W in V is V.

Proof: Since every coset of W in V is a subset of V, the union is certainly a subset of V. Conversely, given $v \in V$, $v = v + 0 \in v + W$ (as $0 \in W$). Thus, every $v \in V$ belongs to some coset of W in V. Hence V is contained in the union of all the cosets of W in V. Hence, the theorem is established.

We may write the statement of Theorem 12 as

$$V = \bigcup_{v \in V} (v + w)$$

A very special property of cosets is given in the following theorem.

Theorem 13: Two cosets $v_1 + W = v_2 + W$ in V are either equal or disjoint. In fact, $v_1 + W = v_2 + W$ iff $v_1 - v_2 \in W$, for $v_1, v_2 \in V$.

Proof: We have to prove that, for $v_1, v_2 \in V$ either $(v_1 + W) \cap (v_2 + W) = \{0\}$ or $v_1 + W = v_2 + W$. Now, suppose $(v_1 + W) \cap (v_2 + W) \neq \{0\}$. Then they have a common non-zero element v, say. That is, $v = v_1 + w_1 = v_2 + w_2$, for some $w_1, w_2 \in W$.

Then $v_1 - v_2 = w_2 - w_1 \in W$ .... (1)

We want to prove that $v_1 + W = v_2 + W$. For this we prove that $v_1 + W \subseteq v_2 + W$ and $v_2 + W \subseteq v_1 + W$.

Now, $u \in v_1 + W \Rightarrow u = v_1 + w_3$, where $w_3 \in W$

$\Rightarrow u = v_2 + (w_2 - w_1) + w_3$, by (1)

$= v_2 + w_4$, where $w_4 \in W$

$\Rightarrow u \in v_2 + W$.

Hence, $v_1 + W \subseteq v_2 + W$.

We can similarly show that $v_2 + W \subseteq v_1 + W$.

Hence, $v_1 + W = v_2 + W$. Note that we have shown that

$v_1 - v_2 \in W \Longrightarrow v_1 + W = v_2 + W$.

$v_1 + W = v_2 + W \Leftrightarrow v_1 - v_2 \in W$

E E23) In the proof above, we have essentially proved that $v_1 - v_2 \in W \Longrightarrow v_1 + W = v_2 + W$. The converse of this is also true. Prove it.

Note: The last two theorems tell us that if W is a subspace of V, then W partitions V into mutually disjoint subsets (namely, the cosets of W in V).

Consider the following example in which we show how a vector space can be partitioned by cosets of a subspace.

Example 22: Consider the subspace $W = \{\alpha(1,0,0) | \alpha \in \mathbf{R}\}$ of $\mathbf{R}^3$. How can you write $\mathbf{R}^3$ as the union of disjoint cosets of W?

Solution: Note that W is just the x-axis in 3-dimensional space.

Any coset of W is of the form

$(a,b,c) + W = \{(a,b,c) + (\alpha,0,0) | \alpha \in \mathbf{R}\} = \{(a+\alpha, b, c) | \alpha \in \mathbf{R}\}$.

Now, for any $(a,b,c) \in \mathbf{R}^3$, $(a,b,c) - (0,b,c) = (a,0,0) \in W$.

Therefore, $(a,b,c) + W = (0,b,c) + W$. Also, the cosets

$(0,b,c) + W$ and $(0,b',c') + W$ are the same iff $b = b'$ and $c = c'$.

Thus, $\{(0,b,c) + W | b,c \in \mathbf{R}\}$ is the set of disjoint cosets of W in $\mathbf{R}^3$.

And $\mathbf{R}^3 = \bigcup \{(0,b,c) + W | b,c \in \mathbf{R}\}$.

Geometrically, the coset $(0,b,c) + W$ represents a line (in the plane determined by the point $(0,b,c)$ and the x-axis) which is parallel to the x-axis and passes through the point $(0,b,c)$. Thus, $\mathbf{R}^3$ is the union of all such distinct lines.

E24) Write $\mathbf{R}^2$ as a disjoint union of the cosets of

a) the subspace $\{(0,0)\}$, b) the subspace $\mathbf{R}^2$.

Before we proceed, let us stress that our notation for a coset of W in V has a peculiarity. A coset $v_1 + W$ can also be written as $v_2 + W$ provided $v_1 - v_2 \in W$. So the same coset can be written in many different ways. Indeed, if C is a coset of W in V, then $C = v + W$, for any vector v in C.

Let us now see how the set of all cosets of W in V can form a vector space.

3.7.2 The Quotient Space

We have pointed out that generally a coset $v + W$ of a subspace W of a vector space V is not itself a subspace of V. We shall now prove that if we take the set of all cosets of W in V, this set can be made into a vector space by defining addition and scalar multiplication suitably.

Notation: Let W be a subspace of the vector space V. We denote the set of all cosets of W in V by V/W. Thus, $V/W = \{v+W | v \in V\}$.

Consider the following example.

Example 23: Let P be the vector space of real polynomials in x and $W = \{f | f \in P, f(0) = 0\}$ be the subspace of P consisting of all those polynomials whose constant term is zero. Show that $P/W = \{a+W | a \in \mathbf{R}\}$.

Solution: Since $a \in P \; \forall \; a \in \mathbf{R}$, certainly $a + W$ is a coset of W in P. So $a+W \in P/W \; \forall \; a \in \mathbf{R}$. Conversely, take an element of P/W, say $f(x) + W$, where f(x) is a polynomial. Suppose

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0, \; a_i \in \mathbf{R}.$$

Then $f(x) = a_0 + g(x)$, where $g(x) = a_1x + a_2x^2 + \dots + a_nx^n$.

Since $g(0) = 0$, $g \in W$.

Hence, $f = a_0 + g$, where $g \in W$.

Thus, $f+W = a_0 + W$ (Theorem 13).

Hence, $f+W \in \{a+W | a \in \mathbf{R}\}$.

This completes the proof that $P/W = \{a+W | a \in \mathbf{R}\}$.

E25) If P_n denotes the vector space of all polynomials of degree $\leq n$, prove that $P_3/P_2 = \{ax^3 + P_2 | a \in \mathbf{R}\}$.

(Hint : For any $f(x) \in P_3$, $\exists \; a \in \mathbf{R}$ such that $f(x) - ax^3 \in P_2$.)

We now proceed to introduce two operations on the set V/W, namely, addition and scalar multiplication.

Definition: Let W be a subspace of V. We define addition on V/W by
$(v_1+W) + (v_2+W) = (v_1+v_2) + W$.

If $\alpha \in R$, $v+W \in V/W$, then we define scalar multiplication on V/W by
$\alpha\cdot(v+W) = (\alpha v) + W$.

Note that our definitions of addition and scalar multiplication seem to depend on the way in which we write a coset. Let us explain this. Suppose C_1 and C_2 are two cosets. What is $C_1 + C_2$? To find $C_1 + C_2$ we must express C_1 as $v_1 + W$ and C_2 as $v_2 + W$. Having done this we can then say that

$C_1 + C_2 = (v_1 + v_2) + W$.

But C_1 can be written in the form $v + W$ in many ways and the same is true for C_2. So the question arises : Is $C_1 + C_2$ dependent on the particular way of writing C_1 and C_2, or is it independent of it? In other words, suppose $C_1 = v_1 + W = v_1' + W$ and $C_2 = v_2+W = v_2'+W$. Then we may say that
$C_1+C_2 = (v_1+W) + (v_2+W) = (v_1+v_2) + W$; but we may also say that
$C_1 + C_2 = (v_1'+W) + (v_2'+ W) = (v_1' + v_2') + W$.

Are these two answers the same? If they are not, which one is to be C_1+C_2? A similar question can arise in the case of αC where α is a scalar and C a coset. These are important questions. Fortunately, they have simple answers as shown by the following theorem.

Theorem 14: Let W be a subspace of a vector space V. If $v_1+W = v_1'+W$ and $v_2 + W = v_2' + W$, then
a) $(v_1+v_2) + W = (v_1' + v_2') + W$

Also, if α is any scalar, then
b) $(\alpha v_1) + W = (\alpha v_1') + W$

Proof: a) For $v_1, v_1', v_2, v_2' \in V$, $v_1 + W = v_1' + W$, $v_2 + W = v_2' + W$
$\Rightarrow v_1 - v_1' \in W$, $v_2 - v_2' \in W$ (by E 23)
$\Rightarrow (v_1 - v_1') + (v_2 - v_2') \in W$
$\Rightarrow (v_1 + v_2) - (v_1'+v_2') \in W$
$\Rightarrow (v_1+v_2) + W = (v_1' + v_2') + W$ (by Theorem 13).
Thus, (a) is true.

b) For any scalar α and $v_1, v_1' \in V$, $v_1 + W = v_1' + W \Rightarrow v_1-v_1' \in W$
$\Rightarrow \alpha (v_1 - v_1') \in W$
$\Rightarrow \alpha v_1 - \alpha v_1' \in W$
$\Rightarrow \alpha v_1 + W = \alpha v_1' + W$
Thus (b) is also proved.

Theorem 14 assures us that the sum and scalar multiplication of cosets is independent of the particular way in which a coset is written. We express this by saying that addition and scalar multiplication of cosets are well defined by the way we have defined them.

This also means that when adding two cosets or when multiplying a scalar and a coset we are free to use any representation for the cosets involved.

We now come to the actual proof that V/W is a vector space.

Theorem 15: Let V be a vector space over a field F, and W be a subspace. Then V/W is a vector space over F.

Proof: We will show that VS1 – VS10 hold for V/W where the operations are addition and scalar multiplication as defined above.
i) VS1 is true since the sum of two cosets is a coset.

ii) For v_1, v_2, v_3 in V we know that $(v_1 + v_2) + v_3 = v_1 + (v_2+v_3)$.
Therefore,

$\{(v_1+W)+(v_2+W)\} + (v_3 + W) = \{(v_1+v_2) + W\} + (v_3 + W)$
$= \{(v_1 + v_2) + v_3\} + W = \{v_1 + (v_2 + v_3)\} + W$
$= (v_1 + W) + \{(v_2 + v_3) + W\}$
$= (v_1 + W) + \{(v_2 + W) + (v_3 + W)\}$

Thus, VS2 is true.

iii) We claim that the coset $0 + W = W$ (since $0 \in W$) is the identity element for V/W.
For $v \in V$, $W + (v + W) = (0+W) + (v+W) = (0+v)+W = v+W$.
Similarly, $(v+W) + W = (v+W) + (0+W) = v+W$. Hence W is the 'zero' of V/W, and VS3 is true.

iv) The additive inverse of $v + W$ is $(-v) + W$, because
$(v+W) + \{(-v) + W\} = (v+(-v)) + W = 0 + W = W$ and
$(-v) + W + (v+W) = (-v+v) + W = 0 + W = W$. This proves that VS4 is true.

v) We note that addition in V is already commutative because V is a vector space. So
$\forall v_1, v_2 \in V, v_1 + v_2 = v_2 + v_1$.
Hence $(v_1+W) + (v_2+W) = (v_1+v_2) + W$
$= (v_2 + v_1) + W$
$= (v_2 + W) + (v_1 + W)$

Thus VS5 holds for V/W.

vi) VS6 is true because, for $\alpha \in F$ and $v + W \in V/W$, $\alpha(v+W) = \alpha v + W \in V/W$.

vii) To prove that VS7 holds, let $\alpha \in F$ and $u, v \in V$. Then
$\alpha\{(u+W) + (v+W)\} = \alpha\{(u+v)+W\}$
$= \alpha(u+v) + W$
$= (\alpha u+\alpha v) + W$
$= (\alpha u+W) + (\alpha v+W)$
$= \alpha(u+W) + \alpha(v+W)$.

viii) For any $\alpha, \beta \in F$ and $v \in V$ you can show, as above, that $(\alpha+\beta)(v+W) = \alpha(v+W) + \beta(v+W)$. Thus, VS8 holds.

ix) For any $\alpha, \beta \in F$ and $v \in V$ we have
$\alpha(\beta(u+W)) = \alpha(\beta u + W)$
$= (\alpha\beta) u + W$
$= (\alpha\beta)(u+W)$

Thus VS9 is true for V/W.

x) For $u \in V$, we have $1.(u+W) = (1 \cdot u) + W = u + W$.
Thus, VS10 holds for V/W.

The vector space we have just obtained has a name.

Definition: If W is a subspace of V, then the vector space V/W is called the **quotient space of V by W.**

The name quotient is very apt because, in a sense, we quotient out the elements of W from those of V.

Example 24: Let V be a vector space over F and $W = \{0\}$. What is V/W?

Solution: $V/W = \{v+W \mid v \in V\} = \{v + \{0\} \mid v \in V\}$
$= \{v \mid v \in V\} = V$

E26) Let $W = \{\alpha(0,1) | \alpha \in R\}$. What is R^2/W?

E27) For any vector space V, show that V/V has only 1 element, namely, the coset V.

And now, let us see what we have done in this unit.

3.8 SUMMARY

Let us conclude the unit by summarising what we have covered in it.

In this unit we have

1) defined a general vector space.
2) given several examples of vector spaces.
3) proved some important properties of a general vector space.
4) defined the notion of a subspace and given criteria to identify subspaces.
5) introduced the idea of the linear span of a set of vectors.
6) shown that the intersection of subspaces of a vector space is a subspace.
7) defined the sum and direct sum of subspaces of a vector space and shown that they are subspaces also.
8) defined cosets and a quotient space.

3.9 SOLUTIONS/ANSWERS

E1) $\forall\ \alpha \in R$ and $(x_1,\ldots,x_n), (y_1,\ldots,y_n) \in R^n$,

$\alpha[(x_1,\ldots,x_n) + (y_1,\ldots,y_n)] = \alpha(x_1+y_1, x_2+y_2,\ldots,x_n+y_n)$

$= (\alpha(x_1+y_1), \alpha(x_2+y_2), \ldots, \alpha(x_n+y_n))$

$= (\alpha x_1 + \alpha y_1, \alpha x_2 + \alpha y_2, \ldots, \alpha x_n + \alpha y_n)$

$= (\alpha x_1, \alpha x_2,\ldots,\alpha x_n) + (\alpha y_1, \alpha y_2,\ldots,\alpha y_n)$

$= \alpha(x_1,x_2,\ldots,x_n) + \alpha(y_1,y_2,\ldots, y_n)$, which proves VS7.

Also, for $\alpha,\beta \in R$, $(\alpha\beta)(x_1,\ldots,x_n) = ((\alpha\beta)x_1, (\alpha\beta)x_2,\ldots, (\alpha\beta)x_n)$

$= (\alpha(\beta x_1), \alpha(\beta x_2),\ldots,\alpha(\beta x_n))$

$= \alpha(\beta x_1, \beta x_2,\ldots,\beta x_n) = \alpha\{\beta(x_1,\ldots,x_n)\}$, which proves VS9.

Finally, $1\cdot(x_1,\ldots,x_n) = (1\cdot x_1, 1\cdot x_2,\ldots,1\cdot x_n) = (x_1,\ldots,x_n)$, which proves VS 10.

E2) For any $\alpha \in R$ and $f \in S$, αf is a function from R to R. Thus, VS6 is satisfied.
To show that VS7 – VS10 are satisfied take any $\alpha, \beta \in R$ and $f, g \in S$. Then, for any $x \in R$, $[\alpha(f+g)](x) = \alpha(f+g)(x) = \alpha(f(x)+g(x)) = \alpha f(x) + \alpha g(x) = (\alpha f)(x) + (\alpha g)(x) = (\alpha f + \alpha g)(x)$
Therefore, $\alpha(f+g) = \alpha f + \alpha g$, that is, VS7 is true.

You can similarly show that $(\alpha+\beta) f = \alpha f + \beta f$, $(\alpha\beta) f = \alpha(\beta f)$ and $1 \cdot f = f$, thus showing that VS8 – VS10 are also true.

E3) Since $(x+ix) + (y+iy) = (x+y)+i(x+y) \in V$, and $\alpha(x+ix) = (\alpha x)+i(\alpha x)$ $\forall \alpha \in R$, and $\forall (x+ix), (y+iy) \in V$, we see that VS1 and VS6 are true. VS2 and VS5 follow from the same properties in R. $0 = 0+i0$ is the additive identity for V, and $(-x) + i(-x)$ is the additive inverse of $x + ix$, $x \in R$.

Also, for any $\alpha, \beta \in R$, and $(x+ix)$, $(y+iy)$ in V the properties VS7 – VS10 can be easily shown to be true. Thus VS1 – VS10 are all true for V.

E4) Addition is a binary operation on Q, since $(ax^2+bx+c) + (dx^2+ex+f) = (a+d)x^2 + (b+e)x + (c+f)$, $\forall a,b,c,d,e,f \in C$.
Scalar multiplication from $C \times Q$ to Q is also well defined since $\alpha(ax^2+bx+c) = \alpha ax^2 + \alpha bx + \alpha c$ $\forall \alpha, a, b, c \in C$. Now on the lines of Example 4, you can show that Q is a complex vector space.

E5) Note that Q′ is a subset of Q in E4. Addition is closed on Q, but not on Q′, because, for example, $2x^2 \in Q'$ and $(-2)x^2 \in Q'$, but $2x^2 + (-2)x^2 \notin Q'$. Thus, Q′ can't be a vector space under the usual operations.

E6) Now $(x,y,z), (x_1,y_1,z_1) \in V$

$\Rightarrow ax + by + cz = 0$ and $ax_1 + by_1 + cz_1 = 0$.

$\Rightarrow a(x+x_1) + b(y+y_1) + c(z+z_1) = 0$.

$\Rightarrow (x+x_1, y+y_1, z+z_1) \in V \Rightarrow$ VS1 is true for V.

Also, for $\alpha \in R$ and $(x,y,z) \in V$, $\alpha(x,y,z) = (\alpha x, \alpha y, \alpha z) \in V$.
This is because $ax + by + cz = 0 \Rightarrow \alpha(ax + by + cz) = 0$.

$\Rightarrow a(\alpha x) + b(\alpha y) + c(\alpha z) = 0$. Thus, VS6 is true for V.

$(0,0,0) \in V$ and is the additive identity for V. Thus, VS3 is true.

For $(x,y,z) \in V$, $(-x,-y,-z) \in V$, and is the additive inverse of (x,y,z). Thus, VS4 is true. VS2 and VS5 are true for V, since they are true for R^3. VS7 – VS9 are true for V, since they are true for R^3. VS10 is true, by definition of scalar multiplication.

E7) $C^n = \{(x_1, x_2, \ldots, x_n) | x_i \in C\}$. This problem can be solved on the lines of Example 2.

E8) From Theorem 1 you know that $(-\alpha)u = -(\alpha u)$ $\forall \alpha \in F$. In particular, $(-1)u = -u$.
Therefore, $(-1)(-u) = (-1)[(-1)u]$

$= [(-1)(-1)]u$, by VS9.

$= 1 \cdot u = u$.

E9) Now, $(u+v) + (-u-v) = (v+u) + (-u-v)$, by VS5

$= [v + (u + (-u))] + (-v)$, by VS2

$= (v+0) + (-v) = v + (-v) = 0$

Thus, by VS4, $-(u+v) = -u - v$.

E10) $-(-u) = (-1)(-u)$ by Theorem 1.

$= u$, by E8.

E11) $\alpha(u-v) = \alpha(u+(-v)) = \alpha u + \alpha(-v) = \alpha u + \alpha(-1)v = \alpha u + (-\alpha)v$
$= \alpha u - \alpha v$.

E12) This is a particular case of the vector space in E6 (with $a = 2, b = 1, c = -1$).
If $\alpha, \beta \in \mathbf{R}$ and $(x_1, x_2, x_3), (y_1, y_2, y_3) \in W$, then $\alpha(x_1,x_2,x_3) + \beta(y_1,y_2,y_3) =$
$(\alpha x_1+\beta y_1, \alpha x_2+\beta y_2, \alpha x_3+\beta y_3)$. Also,
$2(\alpha x_1 + \beta y_1) + (\alpha x_2 + \beta y_2) - (\alpha x_3+\beta y_3)$
$= \alpha (2x_1+x_2-x_3) + \beta(2y_1+y_2-y_3) = 0$, since
$2x_1 + x_2 - x_3 = 0$ and $2y_1 + y_2 - y_3 = 0$.
Thus, $\alpha (x_1,x_2,x_3) + \beta(y_1,y_2,y_3) \in W$.
Hence, W is a subspace of $\mathbf{R}^3$.

E13) a) $W = \{(x_1, -x_1, x_2) | x_1, x_2 \in R\}$. $W \neq \phi$, since $(0,0,0) \in W$.
For $\alpha,\beta \in R$ and $(x_1,-x_1,x_2), (y_1,-y_1,y_2) \in W$, we have $\alpha(x_1,-x_1,x_2) +$
$\beta (y_1,-y_1,y_2) = (\alpha x_1+\beta y_1, -(\alpha x_1+\beta y_1), \alpha x_2+\beta y_2) \in W$. ∴, W is a vector space

b) $W = \{(x_1,x_2,x_3) \in \mathbf{R}^3 | x_1^2 \geq 0\}$
Since $x_1^2 \geq 0 \ \forall \ x_1 \in \mathbf{R}$, we see that $W = \mathbf{R}^3$, and hence is a vector space.

c) $W = \{(x_1,x_2,x_3) \in \mathbf{R}^3 | x_1x_2 = 0\}$.
$W \neq \phi$, since $(0,0,0) \in W$.
Now, $(1,0,0) \in W$ and $(0,1,0) \in W$, but $(1,0,0) + (0,1,0) = (1,1,0) \notin W$.
∴, W is not a subspace of $\mathbf{R}^3$.

d) $W = \{(x_1,x_2,x_3) \in \mathbf{R}^3 | x_1+x_2+x_3 = 1\}$.
Now, $(1,0,0)$ and $(0,1,0) \in W$, but $(1,0,0) + (0,1,0) = (1,1,0) \notin W$,
since $1+1+0 = 2 \neq 1$. ∴, W is not a subspace of $\mathbf{R}^3$.

E14) Firstly, $\{0\}$ is non-empty. Secondly, $0 + 0 = 0 \in \{0\}$ and $\alpha \cdot 0 = 0 \in \{0\}$, for any $\alpha \in F$.
Thus, by Theorem 3, $\{0\}$ is a subspace of V.

E15) $[S] = \{a+bx+cx^2 | a,b,c \in R\}$. ∴, $2x+3x^3 \notin [S]$.

E16) $[S] = \{\alpha(1,2,1) + \beta(2,1,0) | \alpha,\beta \in R\}$
$= \{(\alpha+2\beta, 2\alpha+\beta, \alpha) | \alpha,\beta \in R\}$

a) $(5,3,1) \in [S] \Leftrightarrow \exists \ \alpha,\beta \in R$ such that $\alpha+2\beta = 5, 2\alpha+\beta = 3, \alpha = 1$.
Now, $\alpha=1$ and $\alpha+2\beta = 5 \Rightarrow \beta = 2$. But then $2\alpha+\beta = 2 + 2 = 4 \neq 3$.
∴ $(5,3,1) \notin [S]$.

b) $(2,1,0) \in S \subseteq [S]$. ∴, $(2,1,0) \in [S]$

c) $(4,5,2) \in [S]$

E17) $[S] = \{ax + b(x^2+1) + c(x^3-1) | a,b,c \in R\}$
$= \{cx^3 + bx^2 + ax + (b-c) | a,b,c \in R\}$
a) $x^2 + x + 1 \in [S]$, taking $a = 1, b = 1, c = 0$
b) $2x^3 + x^2 + 3x + 2 \notin [S]$, since $b = 1, c = 2$ and $b-c \neq 2$.

E18) If $(x,y,z) \in U \cap W$ then $(x,y,z) \in U$ and $(x,y,z) \in W$.
Now, $(x,y,z) \in U \Rightarrow z = 2x$, and
$(x,y,z) \in W \Rightarrow y = 2x$
∴, Any element of $U \cap W$ is of the form $(x,2x,2x)$, $x \in R$. That is, $U \cap W =$
$\{(x,2x,2x) | x \in R\}$.

E19) $[A \cup B] = A + B = \{(x,0,0) + (0,y,0) | x,y \in R\}$
$= \{(x,y,0) | x,y \in R\}$

E20) Each of A + C and B + C are subspaces of R^3. Now, for any $(a,b,c) \in R^3$,

$(a,b,c) = (a,b,-a-b) + (0,0,a+b+c) \in A + C$, and,

$(a,b,c) = (a,b,a) + (0,0,c-a) \in B + C$.

Therefore, $R^3 = A+C$ and $R^3 = B+C$

Now $A \cap C = \{(x,y,z) \in R^3 | x+y+z = 0 \text{ and } x = 0 = y\}$

$= \{(0,0,0)\}$, $\therefore$ A + C is a direct sum.

Also $B \cap C = \{(x,y,z) \in R^3 | x = z \text{ and } x = 0 = y\}$

$= \{(0,0,0)\}$, $\therefore$ B+C is also a direct sum.

E21) Firstly, $A+B \subseteq C$.

Secondly, $A \cap B = \{x+iy | y = 0 \text{ and } x = 0\} = \{0\}$. This means that the sum, A+B, is a direct sum. Finally, take any element $x + iy \in C$.

Then $x + iy = (x+i0) + iy \in A + B$

Therefore, $C = A \oplus B$.

E22) a) Since $v \in W$, $v + W = W$

b) $v + W = \{(x-2)(x^2+1) + f(x) | f(x) \in P \text{ and } f(1) = 0\}$

E23) $v_1 + W = v_2 + W \Rightarrow v_1 \in v_1 + W = v_2 + W$

$\Rightarrow v_1 \in v_2 + W \Rightarrow v_1 = v_2 + w$, for some $w \in W$

$\Rightarrow v_1 - v_2 = w \in W \Rightarrow v_1 - v_2 \in W$.

E24) a) Any coset of $\{(0,0)\}$ in R^2 is $(a,b) + \{(0,0)\} = \{(a,b)\}$. Thus two cosets (a,b) and (c,d) are disjoint, iff $(a,b) \neq (c,d)$, i.e., iff (a,b) and (c,d) are distinct elements of R^2. Thus, $R^2 = \cup\{(a,b) + \{(0,0)\} \mid a,b \in R\} = \cup\{(a,b) \mid a,b \in R\}$

b) Any coset $(a,b) + R^2 = R^2$, since $(a,b) \in R^2$. Thus, the only coset of R^2 in R^2 is R^2 itself. So the disjoint union is only R^2.

E25) $P_3/P_2 = \{(ax^3+bx^2+cx+d) + P_2 | a,b,c,d \in R\}$.

Now, $\{ax^3+P_2 | a \in R\} \subseteq P_3/P_2$. Conversely, any element of P_3/P_2 is $(ax^3 + bx^2+cx+d) + P_2$, where $a,b,c,d \in R$. Now $(ax^3+bx^2+cx+d) - ax^3 = bx^2+cx+d \in P_2$.

Therefore, $(ax^3+bx^2+cx+d) + P_2 = ax^3 + P_2$ (by Theorem 13)

$\in \{ax^3 + P_2 | a \in R\}$.

Thus, $P_3/P_2 = \{ax^3 + P_2 | a \in R\}$.

E26) Firstly, note that W is a subspace of R^2, and hence R^2/W is meaningful. Now $R^2/W = \{(a,b) + W | a,b \in R\}$

For any $(a,b) \in R^2$, we have.

$(a,b) - (a,0) = (0,b) \in W$

$\therefore$, $(a,b) + W = (a,0) + W$

Therefore, $R^2/W = \{(a,0) + W | a \in R\}$

E27) $V/V = \{v+V | v \in V\}$. But $v + V = V \ \forall \ v \in V$. $\therefore$, $V/V = V$.

UNIT 4 BASIS AND DIMENSION

Structure

4.1 INTRODUCTION

In the last unit you saw that the linear span [S] of a non-empty subset S of a vector space V is the smallest subspace of V containing S. In this unit we shall consider the question of finding a subset S of V such that S generates the whole of V, i.e., [S] = V. Of course, one such subset of V is V itself, as [V] = V. But there also are smaller subsets of V which span V. For example, if $S = V \setminus \{0\}$, then [S] contains 0, being a vector space. [S] also contains S. Thus, it is clear that [S] = V. We therefore ask: What is the smallest (minimal) subset B of V such that [B] = V? That is, we are looking for a subset B of V which generates V and, if we take any proper subset C of B, then $[C] \neq V$. Such a subset is called a basis of V.

We shall see that if V has one basis B, which is a finite set, then all the bases of V are finite and all the bases have the same number of elements. This number is called the dimension of the vector space. We shall also consider relations between the dimensions of various types of vector spaces.

As in the case of previous units, we suggest that you go through this unit very carefully because we will use the concepts of 'basis' and 'dimension' again and again.

Objectives

After studying this unit, you should be able to

- decide whether a given set of vectors in a vector space is linearly independent or not;
- determine whether a given subset of a vector space is a basis of the vector space or not;
- construct a basis of a finite-dimensional vector space;
- obtain and use formulae for the dimensions of the sum of two subspaces, intersection of two subspaces and quotient spaces.

4.2 LINEAR INDEPENDENCE

In Section 3.5 we discussed the concept of a linear combination of vectors. Look closely at the following two subsets of P, the real vector space of all polynomials, namely, $S_1 = \{1, x, x^2, \ldots\ldots, x^n\}$ and $S_2 = \{1, x^2, x^2 + 5\}$. Consider any linear combination of elements of S_1 : $\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots\ldots + \alpha_n x^n$, where $\alpha_i \in R \; \forall \; i = 0, 1, \ldots\ldots, n$. This sum is equal to zero if and only if each of the α_i's is zero. On the other hand, consider the linear combination of elements of S_2: $\beta_0 + \beta_1 x^2 + \beta_2(x^2 + 5)$, where $\beta_0 = 5$, $\beta_1 = 1$, $\beta_2 = -1$. This sum is zero. What we have just seen is that the elements of S_1 are linearly independent, while those of S_2 are linearly dependent. To understand what this means let us consider the following definitions.

Definition: If V is a vector space over a field **F**, and if $v_1, \ldots, v_n$ are in V, we say that they are **linearly dependent** over **F** if there exist elements $\alpha_1, \ldots, \alpha_n$ in **F** such that $\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$, with $\alpha_i \neq 0$ for some i.
If the vectors $v_1, \ldots, v_n$ are not linearly dependent over **F**, they are said to be **linearly independent** over **F**.

Note: For convenience, we contract the phrase 'linearly independent (or dependent) over **F**' to 'linearly independent (or dependent)' if there is no confusion about the field we are working with.

Note that linear independence and linear dependence are mutually exclusive properties, i.e., no set can be both linearly independent and linearly dependent. It is also clear that any set of n vectors in a vector space is either linearly independent or linearly dependent.

You must remember that, even for a linearly independent set $v_1, \ldots\ldots, v_n$, there is a linear combination
$0.v_1 + 0.v_2 + \ldots\ldots + 0.v_n = 0$, in which all the scalars are zero. In fact this is the only way that zero can be written as a linear combination of linearly independent vectors.

We are, therefore, led to assert the following criterion for linear independence :

A set, $v_1, v_2, \ldots, v_n$ is linearly independent iff
$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0 \Rightarrow \alpha_i = 0 \ \forall \ i.$

We will often use this criterion to establish the linear independence of a set.

Thus, to check whether $v_1, \ldots, v_n$ is linearly independent or linearly dependent, we usually proceed as follows:

1) Assume that $\sum_{i=1}^{n} \alpha_i v_i = 0$, α_i scalars.

2) Try to prove that each $\alpha_i = 0$.

If this can be proved, we can conclude that the given set is linearly independent. But if, on the other hand, we can find $\alpha_1, \alpha_2, \ldots, \alpha_n$ not all zero, such that $\sum_{i=1}^{n} \alpha_i v_i = 0$, then we must conclude that $v_1, \ldots, v_n$ is linearly dependent.

Consider the following examples.

Example 1: Check whether the following subsets of $\mathbf{R}^3$ or $\mathbf{R}^4$ (as the case may be) are linearly independent or not.

a) $\{u = (1, 0, 0), v = (0, 0, -5)\}$
b) $\{u = (-1, 6, -12), v = (\frac{1}{2}, -3, 6)\}$
c) $\{u = (1,2,3,4), v = (4,3,2,1)\}$

Solution: a) Let $a u + b v = 0$, $a, b \in \mathbf{R}$.

Then, $a(1,0,0) + b(0,0,-5) = (0,0,0)$
i.e., $(a,0,0) + (0,0,-5b) = (0,0,0)$
i.e., $(a, 0, -5b) = (0,0,0)$
i.e., $a = 0, -5b = 0$, i.e., $a = 0, b = 0$.
$\therefore$, $\{u,v\}$ is linearly independent.

b) Let $a u + b v = 0$, $a, b \in \mathbf{R}$.

Then $(-a, 6a, -12a) + (\frac{b}{2}, -3b, 6b) = (0, 0, 0)$
i.e., $-a + \frac{b}{2} = 0$, $6a - 3b = 0$, $-12a + 6b = 0$. Each of these equations is equivalent to $2a - b = 0$, which is satisfied by many non-zero values of a and b (e.g., $a = 1, b = 2$). Hence, $\{u,v\}$ is linearly dependent.

c) Suppose $a\,u + b\,v = 0$, $a, b \in \mathbf{R}$. Then

$(a + 4b, 2a + 3b, 3a + 2b, 4a + b) = (0,0,0,0)$

i.e., $a + 4b = 0$ .... (1)

$2a + 3b = 0$ .... (2)

$3a + 2b = 0$ .... (3)

$4a + b = 0$ .... (4)

Subtracting (2) from (3) we get $a - b = 0$, i.e., $a = b$. Putting this in (1), we have $5\,b = 0$. $\therefore$, $b = 0$, and so, $a = b = 0$. Hence, $\{u,v\}$ is linearly independent.

Example 2: In the real vector space of all functions from R to R, determine whether the set $\{\sin x, e^x\}$ is linearly independent.

Solution: The zero element of this vector space is the zero function, i.e., it is the function 0 such that $0(x) = 0 \; \forall \; x \in \mathbf{R}$. So we have to determine $a, b \in \mathbf{R}$ such that, $\forall \; x \in \mathbf{R}$, $a \sin x + be^x = 0$.

In particular, putting $x = 0$, we get $a.0 + b.1 = 0$, i.e., $b = 0$. So our equation reduces to $a \sin x = 0$. Then putting $x = \pi/2$, we have $a = 0$. Thus, $a = 0$, $b = 0$. So, $\{\sin x, e^x\}$ is linearly independent.

You know that the set $\{1, x, x^2, \ldots\ldots, x^n\} \subseteq \mathbf{P}$ is linearly independent. For larger and larger n, this set becomes a larger and larger linearly independent subset of P. This example shows that in the vector space P, we can have as large a linearly independent set as we wish. In contrast to this situation look at the following example, in which more than two vectors are not linearly independent.

Example 3: Prove that in $\mathbf{R}^2$ any three vectors from a linearly dependent set.

Solution: Let $u = (a_1, a_2)$, $v = (b_1, b_2)$, $w = (c_1, c_2) \in \mathbf{R}^2$. If any of these is the zero vector, say $u = (0,0)$, then the linear combination $1.u + 0.v + 0.w$, of u,v,w, is the zero vector, showing that the set $\{u,v,w\}$ is linearly dependent. Therefore, we may suppose that u,v,w, are all non-zero.

We wish to prove that there are real numbers α, β, τ, not all zero, such that $\alpha u + \beta v + \tau w = 0$. That is, $\alpha u + \beta v = -\tau w$. This reduces to the pair of equations.

$\alpha a_1 + \beta b_1 = -\tau c_1$

$\alpha a_2 + \beta b_2 = -\tau c_2$

We can solve this pair of equations to get values of α, β in terms of $a_1, a_2, b_1, b_2, c_1, c_2$ and τ iff $a_1b_2 - a_2b_1 \neq 0$. So, if

$$a_1b_2 - a_2b_1 \neq 0, \text{ we get } \alpha = \frac{\tau(b_1c_2 - b_2c_1)}{a_1b_2 - a_2b_1},$$

$$\beta = \frac{\tau(c_1a_2 - a_2c_1)}{a_1b_2 - a_2b_1}.$$

Then, we can give τ a non-zero value and get the corresponding values of α and β. Thus, if $a_1b_2 - a_2b_1 \neq 0$ we see that $\{u,v,w\}$ is a linearly dependent set.

Suppose, $a_1b_2 - a_2b_1 = 0$. Then one of a_1 and a_2 is non-zero since $u \neq 0$. Similarly, one of b_1 and $b_2 \neq 0$. Let us suppose that $a_1 \neq 0$, $b_1 \neq 0$. Then, observe that

$b_1\,(a_1, a_2) - a_1\,(b_1, b_2)$

$= (b_1a_1, b_1a_2) - (a_1b_1, a_1b_2)$

$= (0,0)$

i.e., $b_1u - a_1v + 0.w = 0$ and, $a_1 \neq 0$, $b_1 \neq 0$.

Hence, in this case also $\{u,v,w\}$ is a linearly dependent set.

Try the following exercises now.

E E1) Check whether each of the following subsets of R^3 is linearly independent.

a) $\{(1,2,3), (2,3,1), (3,1,2)\}$

b) $\{(1,2,3), (2,3,1), (-3,-4,1)\}$

c) $\{(-2,7,0), (4,17,2), (5,-2,1)\}$

d) $\{(-2,7,0), (4,17,2)\}$

E2) Prove that in the vector space of all functions from $\mathbb{R}$ to $\mathbb{R}$, the set $\{\sin x, \cos x\}$ is linearly independent, and the set $\{\sin x, \cos x, \sin(x + \pi/6)\}$ is linearly dependent.

E3) Determine whether each of the following subsets of P is linearly independent or not.

a) $\{x^2, x^2 + 1\}$

b) $\{x^2 + 1, x^2 + 11, 2x^2 - 3\}$

c) $\{3, x + 1, x^2, x^2 + 2x + 5\}$

d) $\{1, x^2, x^3, x^2 + 1\}$

Let us now look more closely at the concept of linear independence.

4.3 SOME ELEMENTARY RESULTS

In this section we shall study some simple consequences of the definition of linear independence. An immediate consequence is the following theorem.

Theorem 1: If $0 \in \{v_1, v_2, \ldots, v_n\}$, a subset of the vector space V, then the set $\{v_1, v_2, \ldots, v_n\}$ is linearly dependent.

Proof: 0 is one of the v_i's. We may assume that $v_1 = 0$. Then $1.v_1 + 0.v_2 + 0.v_3 + \ldots + 0.v_n = 0 + 0 + \ldots + 0 = 0$. That is, 0 is a linear combination of $v_1, v_2, \ldots, v_n$ in which all the scalars are not zero. Thus, the set is linearly dependent.

Try to prove the following result yourself.

E4) Show that, if v is a non-zero element of a vector space V over a field F, then $\{v\}$ is linearly independent.

The next result is also very elementary.

Theorem 2: a) If S is a linearly dependent subset of a vector space V over F, then any subset of V containing S is linearly dependent.

b) A subset of a linearly independent set is linearly independent.

Proof: a) Suppose $S = \{u_1, u_2, \ldots, u_k\}$ and $S \subseteq T \subseteq V$. We want to show that T is linearly dependent.

If S = T there is nothing to prove. Otherwise, let $T = S \cup \{v_1, \ldots, v_m\}$

$= \{u_1, u_2, \ldots, u_k, v_1, \ldots, v_m\}$, where $m > 0$

Now S is linearly dependent. Therefore, for some scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$, not all zero, we have

$$\sum_{i=1}^{k} \alpha_i u_i = 0$$

But then,

$\alpha_1 u_1 + \alpha_2 u_2 + \ldots + \alpha_k u_k + 0.v_1 + 0.v_2 + \ldots + 0.v_m = 0$, with some $\alpha_i \neq 0$. Thus, T is linearly dependent.

b) Suppose $T \subseteq V$ is linearly independent, and $S \subseteq T$. If possible, suppose S is not linearly independent. Then S is linearly dependent, but then by (a), T is also linearly dependent, since $S \subseteq T$. This is a contradiction. Hence, our supposition is wrong. That is, S is linearly independent.

Now, what happens if one of the vectors in a set can be written as a linear combination of the other vectors in the set? The next theorem states that such a set is linearly dependent.

Theorem 3: Let $S = \{v_1, \ldots, v_n\}$ be a subset of a vector space V over a field F. Then S is linearly dependent if and only if some vector of S is a linear combination of the rest of the vectors of S.

Proof: We have to prove two statements here:

i) If some v_i, say v_1, is a linear combination of $v_2, \ldots, v_n$, then S is linearly dependent.

ii) If S is linearly dependent, then some v_i is a linear combination of the other v_i's

Let us prove (i) now. For this, suppose v_1 is a linear combination of $v_2, \ldots, v_n$,
i.e., $v_1 = \alpha_2 v_2 + \ldots + \alpha_n v_n$

$= \sum_{i=2}^{n} \alpha_i v_i$, where $\alpha_i \in F \ \forall_i$. Then $v_1 - \alpha_2 v_2 - \alpha_3 v_3 - \ldots - \alpha_n v_n = 0$,

which shows that S is linearly dependent.

We now prove (ii), which is the converse of (i). Since S is linearly dependent, there exist $\alpha_i \in F$, not all zero, such that

$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0$.

Since some $\alpha_i \neq 0$, suppose $\alpha_k \neq 0$. Then we have

$\alpha_k v_k = -\alpha_1 v_1 \ldots - \alpha_{k-1} v_{k-1} - \alpha_{k+1} v_{k+1} \ldots - \alpha_n v_n$

Since $\alpha_k \neq 0$, we divide throughout by α_k and get

$$v_k = \left(-\frac{\alpha_1}{\alpha_k}\right) v_1 + \ldots + \left(\frac{-\alpha_n}{\alpha_k}\right) v_n = \sum_{\substack{i=1 \\ i \neq k}}^{n} \beta_i v_i,\ \beta_i = \frac{-\alpha_i}{\alpha_k}.$$

Thus, v_k is a linear combination of $v_1, v_2, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n$.

Theorem 3 can also be stated as: S is linearly dependent if and only if some vector in S is in the linear span of the rest of the vectors of S.

Now, let us look at the situation in R^3 where we know that i, j are linearly independent. Can you immediately prove whether the set {i, j, (3,4,5)} is linearly independent or not? The following theorem will help you to do this.

Theorem 4: If S is linearly independent and $v \notin [S]$, then $S \cup \{v\}$ is linearly independent.

Proof: Let $S = \{v_1, v_2, \ldots, v_n\}$ and $T = S \cup \{v\}$.
If possible, suppose T is linearly dependent, then there exist scalars $\alpha, \alpha_1, \alpha_2, \ldots, \alpha_n$, not all zero, such that

$\alpha v + \alpha_1 v_1 + \ldots + \alpha_n v_n = 0$.

Now, if $\alpha = 0$, this implies that there exist scalars
$\alpha_1, \alpha_2, \ldots, \alpha_n$, not all zero, such that
$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0$.

But that is impossible as S is linearly independent. Hence
$\alpha \neq 0$. But then,

$$v = \frac{(-\alpha_1)}{\alpha} v_1 + \ldots + \frac{(-\alpha_n)}{\alpha} v_n,$$

i.e., v is a linear combination of $v_1, v_2, \ldots, v_n$, i.e., $v \in [S]$, which contradicts our assumption.

Therefore, $T = S \cup \{v\}$ must be linearly independent.

Using this theorem we can immediately see that the set {i, j, (3,4,5)} is linearly independent, since (3,4,5) is not a linear combination of i and j.

Now try the following exercises.

E5) Given a linearly independent subset S of a vector space V, can we always get a set T such that $S \subseteq T$ and T is linearly independent?
(Hint: Consider the real space R^2 and the set $S = \{(1,0), (0,1)\}$.)

If you've done E5 you will have found that, by adding a vector to a linearly independent set, it may not remain linearly independent. Theorem 4 tells us that if, to a linearly independent set, we add a vector which is not in the linear span of the set, then the augmented set will remain linearly independent. Thus, the way of generating larger and larger linearly independent subsets of a non-zero vector space V is as follows:

1) Start with any linearly independent set S_1 of V, for example, $S_1 = \{v_1\}$, where $0 \neq v_1 \in V$.

2) If S_1 generates the whole vector space V, i.e., if $[S_1] = V$, then every $v \in V$ is a linear combination of S_1. So $S_1 \cup \{v\}$ is linearly dependent for every $v \in V$. In this case S_1 is a maximal linearly independent set, that is, no larger set than S_1 is linearly independent.

3) If $[S_1] \neq V$, then there must be a $v_2 \in V$ such that $v_2 \notin S_1$. Then, $S_1 \cup \{v_2\} = \{v_1, v_2\} = S_2$ (say) is linearly independent. In this case, we have found a set larger than S_1 which is linearly independent, namely, S_2.

4) If $[S_2] = V$, the process ends. Otherwise, we can find a still larger set S_3 which is linearly independent. It is clear that, in this way, we either reach a set which generates V or we go on getting larger and larger linearly independent subsets of V.

   So far we have only discussed linearly independent sets S, when S is finite. What happens if S is infinite?

Definition: An **infinite** subset of a vector space V is said to be **linearly independent** if every finite subset of S is linearly independent.

Thus, an infinite set S is linearly independent if, for every finite subset $\{v_1, v_2, \ldots, v_n\}$ of S, $\exists$ scalars $\alpha_1, \ldots, \alpha_n$, such that $\sum_{i=1}^{n} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0 \; \forall \; i$.

Consider the following example.

Example 4: Prove that the infinite subset $S = \{1, x, x^2, \ldots\}$, of the vector space **P** of all real polynomials in x, is linearly independent.

Solution: Take any finite subset T of S. Then $\exists$ non-negative distinct integers $a_1, a_2, \ldots, a_k$, such that

$T = \{x^{a_1}, x^{a_2}, \ldots, x^{a_k}\}$.

Now, suppose

$$\sum_{i=1}^{k} \alpha_i x^{a_i} = 0, \text{ where } \alpha_i \in R \; \forall \; i.$$

In P, 0 is the zero polynomial, all of whose coefficients are zero. $\therefore \alpha_i = 0 \; \forall \; i$. Hence T is linearly independent. As every finite subset of S is linearly independent, so is S.

E6) Prove that $\{1, x+1, x^2+1, x^3+1, \ldots\}$ is a linearly independent subset of the vector space P.

And now to the section in which we answer the question raised in Sec. 4.1.

4.4 BASIS AND DIMENSION

We will now discuss two concepts that go hand-in-hand, namely, the basis of a vector space and the dimension of a vector space.

4.4.1 Basis

In Unit 2 you discovered that any vector in $\mathbf{R}^2$ is a linear combination of the two vectors (1,0) and (0,1). You can also see that $\alpha(1,0) + \beta(0,1) = (0,0)$ implies that $\alpha = 0$ and $\beta = 0$ (where $\alpha, \beta \in R$). What does this mean? It means that $\{(1,0), (0,1)\}$ is a linearly independent subset of $\mathbf{R}^2$, which generates $\mathbf{R}^2$.

Similarly, the vectors **i**, **j**, **k** generate $\mathbf{R}^3$ and are linearly independent.

In fact, we will see that such sets can be found in any vector space. We call such sets a "basis" of the concerned vector space. Look at the following definition.

Definition: A subset B, of a vector space V, is called a basis of V, if

Plural of 'basis' is 'bases'

i) B is linearly independent, and
ii) B generates V, i.e., [B] = V.

Note that (ii) implies that every vector in V is a linear combination of a finite number of vectors from B.

Thus, $B \subseteq V$ is a basis of V if B is linearly independent and every vector of V is a linear combination of a finite number of vectors of B.

You have already seen that $\{i = (1,0), j = (0,1)\}$ is a basis of $\mathbf{R}^2$.

The following example shows that $\mathbf{R}^2$ has more than one basis.

A vector space can have more than one basis.

Example 5: Prove that $B = \{v_1, v_2\}$ is a basis of $\mathbf{R}^2$, where $v_1 = (1,1)$, $v_2 = (-1,1)$.

Solution: Firstly, for $\alpha, \beta \in \mathbf{R}$, $\alpha v_1 + \beta v_2 = 0$

$\Rightarrow (\alpha, \alpha) + (-\beta, \beta) = (0,0) \Rightarrow \alpha - \beta = 0, \alpha + \beta = 0$

$\Rightarrow \alpha = \beta = 0.$

Hence, B is linearly independent.

Secondly, given $(a,b) \in \mathbf{R}^2$, we can write

$$(a,b) = \frac{b+a}{2} v_1 + \frac{b-a}{2} v_2$$

Thus, every vector in R^2 is a linear combination of v_1 and v_2. Hence, B is also a basis of R^2.

Another important characteristic of a basis is that **no proper subset of a basis can generate the whole vector space.** This is brought out in the following example.

Example 6: Prove that {i} is not a basis of R^2.
(Here i = (1,0).)

Solution: By E4, since $i \neq 0$, {i} is linearly independent.

Now, $[\{i\}] = \{\alpha i \mid \alpha \in R\} = \{(\alpha, 0) \mid \alpha \in R\}$

$\therefore$ $(1,1) \notin [\{i\}]$; so $[\{i\}] \neq R^2$.

Thus, {i} is not a basis of R^2.

Note that {i} is a proper subset of the basis {i, j} of R^2.

E E7) Prove that :
a) B = {i,j,k} is a basis of R^3, where i = (1,0,0), j = (0,1,0), k = (0,0,1).

b) B = {u,v,w} is a basis of R^3, where
u = (1,2,0), v = (2,1,0), w = (0,0,1).

E E8) Prove that $\{1, x, x^2, x^3, \ldots\}$ is a basis of the vector space, **P**, of all polynomials over a field **F**.

E E9) Prove that $\{1, x+1, x^2 + 2x\}$ is a basis of the vector space, P_2, of all polynomials of degree less than or equal to 2.

E10) Prove that $\{1, x+1, 3x-1, x^2\}$ is not a basis of the vector space P_2.

We have already mentioned that no proper subset of a basis can generate the whole vector space. We will now prove another important characteristic of a basis, namely, **no linearly independent subset of a vector space can contain more vectors than a basis of the vector space. In other words, a basis contains the maximum possible number of linearly independent vectors.**

Theorem 5: If $B = \{v_1, v_2, \ldots, v_n\}$ is a basis of a vector space V over a field F, and $S = \{w_1, w_2, \ldots, w_m\}$ is a linearly independent subset of V, then $m \le n$.

Proof: Since B is a basis of V and $w_1 \in V$, w_1 is a linear combination of $v_1, v_2, \ldots, v_n$. Hence, by Theorem 3,

$S_1' = \{w_1, v_1, v_2, \ldots, v_n\}$ is linearly dependent. Since $[B] = V$ and $B \subseteq S_1'$, we have $[S_1'] = V$. As w_1 is a linear combination of $v_1, v_2, \ldots, v_n$, we have

$$w_1 = \sum_{i=1}^{n} \alpha_i v_i, \alpha_i \in F \ \forall\ i = 1, \ldots, n.$$

Now, $\alpha_i \neq 0$ for some i. (Because, otherwise $w_1 = 0$. But, as w_1 belongs to a linearly independent set, $w_1 \neq 0$.)

Suppose $\alpha_k \neq 0$. Then we can just reorder the elements of B, so that v_k becomes v_1. This does not change any characteristic of B. It only makes the proof easier to deal with since we can now assume that $\alpha_1 \neq 0$. Then

$$v_1 = \frac{1}{\alpha_1} w_1 - \sum_{i=2}^{n} \frac{\alpha_i}{\alpha_1} v_i,$$

that is, v_1 is a linear combination of $w_1, v_2, v_3, \ldots, v_n$. So, any linear combination of $v_1, v_2, \ldots, v_n$ can also be written as a linear combination of $w_1, v_2, \ldots, v_n$. Thus, if $S_1 = \{w_1, v_2, v_3, \ldots, v_n\}$, then $[S_1] = V$.

Note that we have been able to replace v_1 by w_1 in B in such a way that the new set still generates V. Next, let

$S_2' = \{w_2, w_1, v_2, v_3, \ldots, v_n\}$.

Then, as above, S_2' is linearly dependent and $[S_2'] = V$.

Also $w_2 = \beta_1 w_1 + \beta_2 v_2 + \ldots + \beta_n v_n$, $\beta_i \in F \ \forall\ i = 1, \ldots, n$.

Again, $\beta_i \neq 0$ for some i, since $w_2 \neq 0$. Also, it cannot happen that $\beta_1 \neq 0$ and $\beta_i = 0$ $\forall\ i \ge 2$, since $\{w_1, w_2\}$ is a linearly independent set (by Theorem 2(b)). So $\beta_i \neq 0$ for some $i \ge 2$.

Again, for convenience, we may assume that $\beta_2 \neq 0$. Then

$$v_2 = \frac{1}{\beta_2} w_2 - \frac{\beta_1}{\beta_2} w_1 - \frac{\beta_3}{\beta_2} v_3 - \ldots\ldots - \frac{\beta_n}{\beta_2} v_n.$$

This shows that v_2 is a linear combination of $w_1, w_2, v_3, \ldots\ldots, v_n$. Hence, if $S_2 = \{w_2, w_1, v_3, v_4, \ldots\ldots, v_n\}$, then $[S_2] = V$.

So we have replaced v_1, v_2 in B by w_1, w_2, and the new set generates V. It is clear that we can continue in the same way, replacing v_i by w_i at the ith step.

Now, suppose $n < m$. Then, after n steps, we will have replaced all v_i's by corresponding w_i's and we shall have a set

$S_n = \{w_n, w_{n-1}, \ldots. w_2, w_1\}$ with $[S_n] = V$. But then, this means that $w_{n+1} \in V = [S_n]$, i.e., w_{n+1} is a linear combination of $w_1, w_2, \ldots.., w_n$. This implies that the set $\{w_1, \ldots., w_n, w_{n+1}\}$ is linearly dependent. This contradicts the fact that $\{w_1, w_2, \ldots.., w_m\}$ is linearly independent. Hence, $m \leq n$.

An immediate corollary of Theorem 5 gives us a very quick way of determining whether a given set is a basis of a given vector space or not.

Corollary 1: If $B = \{v_1, v_2, \ldots\ldots, v_n\}$ is a basis of V, **then any set of n linearly independent vectors is a basis of V.**

Proof: If $S = \{w_1, w_2, \ldots., w_n\}$ is a linearly independent subset of V, then, as shown in the proof of Theorem 5, $[S] = V$. As S is linearly independent and $[S] = V$, S is a basis of V.

The following example shows how the corollary is useful.

Example 7: Show that (1,4) and (0,1) form a basis of $\mathbf{R}^2$ over **R**.

Solution: You know that (1,0) and (0,1) form a basis of $\mathbf{R}^2$ over R. Thus, to show that the given set forms a basis, we only have to show that the 2 vectors in it are linearly independent. For this, consider the equation

$\alpha(1,4) + \beta(0,1) = 0$, where $\alpha, \beta \in \mathbf{R}$. Then $(\alpha, 4\alpha + \beta) = (0,0) \Rightarrow \alpha = 0, \beta = 0$.

Thus, the set is linearly independent. Hence, it forms a basis of $\mathbf{R}^2$.

E E11) Let V be a vector space over F, with $\{u,v,w,t\}$ as a basis.

a) Is $\{u, v + w, w + t, t + u\}$ a basis of V?
b) Is $\{u, t\}$ a basis of V?

We now give two results that you must always keep in mind when dealing with vector spaces. They depend on Theorem 5.

Theorem 6: If one basis of a vector space contains n vectors, then all its bases contain n vectors.

Proof: Suppose $B_1 = \{v_1, v_2, \ldots\ldots, v_n\}$ and $B_2 = \{w_1, w_2, \ldots.., w_m\}$ are both bases of V. As B_1 is a basis and B_2 is linearly independent, we have $m \leq n$, by Theorem 5. On the other hand, since B_2 is a basis and B_1 is linearly independent, $n \leq m$. Thus, $m = n$.

Theorem 7: If a basis of a vector space contains n vectors, then any set containing more than n vectors is linearly dependent.

Proof: Let $B_1 = \{v_1, \ldots, v_n\}$ be a basis of V and $B_2 = \{w_1, \ldots, w_{n+1}\}$ be a subset of V. Suppose B_2 is linearly independent. Then, by Corollary 1 of Theorem 5, $\{w_1, \ldots, w_n\}$ is a basis of V. This means that V is generated by $w_1, \ldots, w_n$. Therefore, w_{n+1} is a linear combination of $w_1, \ldots, w_n$. This contradicts our assumption that B_2 is linearly independent. Thus, B_2 must be linearly dependent.

E12) Using Theorem 7, prove that the subset $S = \{1, x+1, x^2, x^3+1, x^3, x^2+6\}$ of P_3, the vector space of all real polynomials of degree ≤ 3, is linearly dependent.

So far we have been saying that "if a vector space has a basis, then ....". Now we state the following theorem (without proof).

Theorem 8: Every non-zero vector space has a basis.

Note : The space $\{0\}$ has no basis.

Let us now look at the scalars in any linear combination of basis vectors.

Coordinates of a vector: You have seen that if $B = \{v_1, \ldots, v_n\}$ is a basis of a vector space V, then every vector of V is a linear combination of the **elements of B. We now** show that this linear combination is unique.

Theorem 9: If $B = \{v_1, v_2, \ldots, v_n\}$ is a basis of the vector space V over a field **F**, then every $v \in V$ can be expressed **uniquely** as a linear combination of $v_1, v_2, \ldots, v_n$.

Proof: Since $[B] = V$ and $v \in V$, v is a linear combination of $\{v_1, v_2, \ldots, v_n\}$. To prove uniqueness, suppose there exist scalars $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$ such that

$v = \alpha_1 v_1 + \ldots + \alpha_n v_n = \beta_1 v_1 + \ldots + \beta_n v_n$.

Then $(\alpha_1 - \beta_1) v_1 + (\alpha_2 - \beta_2) v_2 + \ldots + (\alpha_n - \beta_n) v_n = 0$.

But $\{v_1, v_2, \ldots, v_n\}$ is linearly independent. Therefore,

$\alpha_i - \beta_i = 0 \ \forall\ i$, i.e., $\alpha_i = \beta_i \ \forall\ i$.

This establishes the uniqueness of the linear combination.

This theorem implies that given a basis B of V, for every $v \in V$, there is one and only one way of writing

$$v = \sum_{i=1}^{n} \alpha_i v_i \text{ with } \alpha_i \in \mathbf{F} \ \forall\ i.$$

Definition: Let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis of an n-dimensional vector space V. Let $v \in V$. If the unique expression of v as a linear combination of $v_1, v_2, \ldots, v_n$ is $v = \alpha_1 v_1 + \ldots + \alpha_n v_n$, then $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ are called the **coordinates of v relative to the basis B**, and α_i is called the **i^{th} coordinate of v.**

The coordinates of a vector will depend on the particular basis chosen, as can be seen in the following example.

Example 8: For R^2, consider the two bases

$B_1 = \{(1,0), (0,1)\}$, $B_2 = \{(1,1), (-1,1)\}$ (see Example 5). **Find the coordinates of the** following vectors in R^2 relative to both B_1 and B_2.

(a) (1,2) (b) (0,0) (c) (p,q).

Solution : (a) Now, $(1,2) = 1(1,0) + 2(0,1)$.

Therefore, the coordinates of $(1,2)$ relative to B_1 are $(1,2)$.

Also, $(1,2) = 3/2\,(1,1) + 1/2\,(-1,1)$. Therefore, the coordinates of $(1,2)$ relative to B_2 are $(3/2, 1/2)$.

(b) $(0,0) = 0(1,0) + 0(0,1)$ and $(0,0) = 0(1,1) + 0(-1,1)$.

In this case, the coordinates of $(0,0)$ relative to both B_1 and B_2 are $(0,0)$.

(c) $(p,q) = p(1,0) + q(0,1)$ and

$$(p,q) = \frac{q+p}{2}(1,1) + \frac{q-p}{2}(-1,1)$$

Therefore, the coordinates of (p,q) relative to B_1 are (p,q) and the coordinates of (p,q) relative to B_2 are $\left(\frac{q+p}{2}, \frac{q-p}{2}\right)$.

Note : The basis $B_1 = \{i,j\}$ has the pleasing property that for all vectors $(p,q) \in \mathbf{R}^2$, the coordinates of (p,q) relative to B_1 are (p,q). For this reason B_1 is called the **standard basis** of $\mathbf{R}^2$, and the coordinates of a vector relative to the standard basis are called **standard coordinates** of the vector. In fact, this is the basis we normally use for plotting points in 2-dimensional space.

In general, the basis

$B = \{(1,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,\ldots,0,1)\}$ of $\mathbf{R}^n$ over $\mathbf{R}$ is called the standard basis of $\mathbf{R}^n$.

Example 9: Let V be the vector space of all real polynomials of degree at most 1 in the variable x. Consider the basis $B = \{5, 3x\}$ of V. Find the coordinates relative to B of the following vectors.

(a) $2x + 1$ (b) $3x - 5$ (c) 11 (d) $7x$.

Solution: a) Let $2x+1 = \alpha(5) + \beta(3x) = 3\beta x + 5\alpha$.

Then $3\beta = 2$, $5\alpha = 1$. So, the coordinates of $2x + 1$ relative to B are $(1/5, 2/3)$.

b) $3x - 5 = \alpha(5) + \beta(3x) \Rightarrow \alpha = -1, \beta = 1$. Hence, the answer is $(-1,1)$.

c) $11 = \alpha(5) + \beta(3x) \Rightarrow \alpha = 11/5, \beta = 0$. Thus, the answer is $(11/5, 0)$.

d) $7x = \alpha(5) + \beta(3x) \Rightarrow \alpha = 0, \beta = 7/3$. Thus, the answer is $(0, 7/3)$.

E E13) Find a standard basis for $\mathbf{R}^3$ and for the vector space P_2 of all polynomials of degree ≤ 2.

E E14) For the basis $B = \{(1,2,0), (2,1,0), (0,0,1)\}$ of $\mathbf{R}^3$, find the coordinates of $(-3,5,2)$.

E15) Prove that, for any basis $B = \{v_1, v_2, \ldots, v_n\}$ of a vector space V, the coordinates of 0 are (0,0,0,……0).

E16) For the basis $B = \{3, 2x+1, x^2-2\}$ of the vector space P_2 of all polynomials of degree ≤ 2, find the coordinates of

(a) $6x + 6$ (b) $(x+1)^2$ (c) x^2

E17) For the basis $B = \{u,v\}$ of R^2, the coordinates of (1,0) are (1/2, 1/2) and the coordinates of (2,4) are (3, −1). Find u,v.

We now continue the study of vector spaces by looking into their 'dimension', a concept directly related to the basis of a vector space.

4.4.2 Dimension

So far we have seen that, if a vector space has a basis of n vectors, then every basis has n vectors in it. Thus, given a vector space, the number of elements in its different bases remains constant.

Definition : If a vector space V over the field F has a basis containing n vectors, we say that the **dimension** of V is n. We write $\dim_F V = n$ or, if the underlying field is understood, we write **dim V = n.**

If $V = \{0\}$, it has no basis. We define $\dim 0 = 0$.

If a vector space does not have a finite basis, we say that it is **infinite-dimensional.**

In E8, you have seen that P is infinite-dimensional. Also E9 says that $\dim_R P_2 = 3$. Earlier you have seen that $\dim_R R^2 = 2$ and $\dim_R R^3 = 3$.

In Theorem 8, you read that every non-zero vector space has a basis. The next theorem gives us a helpful criterion for obtaining a basis of a finite-dimensional vector space.

Theorem 10: If there is a subset $S = \{v_1, \ldots, v_n\}$ of a non-empty vector space V such that $[S] = V$, then V is finite-dimensional and S contains a basis of V.

Proof: We may assume that $0 \notin S$ because, if $0 \in S$, then $S \setminus \{0\}$ will still satisfy the conditions of the theorem. If S is linearly independent then, since $[S] = V$, S itself is a basis of V. Therefore, V is finite-dimensional ($\dim V = n$). If S is linearly dependent, then some vector of S is a linear combination of the rest (Theorem 3). We may assume that this vector is v_n. Let $S_1 = \{v_1, v_2, \ldots, v_{n-1}\}$

Since $[S] = V$ and v_n is a linear combination of $v_1, \ldots, v_{n-1}$, $[S_1] = V$.

If S_1 is linearly dependent, we drop, from S_1, that vector which is a linear combination of the rest, and proceed as before. Eventually, we get a linearly independent subset

$S_r = \{v_1, v_2, \ldots, v_{n-r}\}$

of S, such that $[S_r] = V$ (This must happen because $\{v_1\}$ is certainly linearly independent.) So $S_r \subseteq S$ is a basis of V and $\dim V = n - r$.

Example 10: Show that the dimension of R^n is n.

Solution: The set of n vectors

$\{(1,0,0,\ldots,0), (0,1,0,\ldots,0), \ldots, (0,0,0,\ldots,0,1)\}$ spans V and is obviously a basis of R^n.

E E18) Prove that the real vector space C of all complex numbers has dimension 2.

E E19) Prove that the vector space P_n, of all polynomials of degree at most n, has dimension $n + 1$.

We now see how to obtain a basis once we have a linearly independent set.

4.4.3 Completion of a Linearly Independent Set to a Basis

We have seen that in an n-dimensional vector space, a linearly independent subset cannot have more than n vectors (Theorem 7). We now ask: Suppose we have a linearly independent subset S of an n-dimensional vector space V. Further, suppose S has m ($< n$) vectors. Can we add some vectors to S, so that the enlarged set will be a basis of V? In other words, can we extend a linearly independent subset to get a basis? The answer is yes. But, how many vectors would we have to add? Do you remember Corollary 1 of Theorem 5? That gives the answer: $n-m$. Of course, any $(n-m)$ vectors won't do the job. The vectors have to be carefully chosen. That is what the next theorem is about.

Theorem 11: Let $W = \{w_1, w_2, \ldots, w_m\}$ be a linearly independent subset of an n-dimensional vector space V. Suppose $m < n$. Then there exist vectors $v_1, v_2, \ldots, v_{n-m} \in V$ such that $B = \{w_1, w_2, \ldots, w_m, v_1, v_2, \ldots, v_{n-m}\}$ is a basis of V.

Proof: Since $m < n$, W is not a basis of V (Theorem 6). Hence, $[W] \neq V$. Thus, we can find a vector $v_1 \in V$ such that $v_1 \notin [W]$. Therefore, by Theorem 4, $W_1 = W \cup \{v_1\}$ is linearly independent. Now, W_1 contains $m + 1$ vectors. If $m+1 = n$, W_1 is a linearly independent set with n vectors in the n-dimensional space V, so W_1 is a basis of V (Theorem 5, Cor. 1). That is, $\{w_1, \ldots, w_m, v_1\}$ is a basis of V. If $m+1 < n$, then $[W_1] \neq V$, so there is a $v_2 \in V$ such that $v_2 \notin [W_1]$. Then $W_2 = W_1 \cup \{v_2\}$ is linearly independent and contains $m+2$ vectors. So, if $m+2 = n$, then

$$W_2 = W_1 \cup \{v_2\} = W \cup \{v_1, v_2\} = \{w_1, w_2, \ldots, w_m, v_1, v_2\}$$

is a basis of V. If $m+2 < n$, we continue in this fashion. Eventually, when we have adjoined $n-m$ vectors $v_1, v_2, \ldots, v_{n-m}$ to W, we shall get a linearly independent set $B = \{w_1, w_2, \ldots, w_m, v_1, v_2, \ldots, v_{n-m}\}$ containing n vectors, and hence B will be a basis of V.

Let us see how Theorem 11 actually works.

Example 11: Complete the linearly independent subset $S = \{(2,3,1)\}$ of $\mathbf{R}^3$ to a basis of $\mathbf{R}^3$.

Solution: Since $S = \{(2,3,1)\}$,

$$[S] = \{\alpha(2,3,1) \mid \alpha \in \mathbf{R}\}$$
$$= \{(2\alpha, 3\alpha, \alpha) \mid \alpha \in \mathbf{R}\}$$

Now we have to find $v_1 \in \mathbf{R}^3$ such that $v_1 \notin [S]$, i.e., such that $v_1 \neq (2\alpha, 3\alpha, \alpha)$ for any $\alpha \in \mathbf{R}$. We can take $v_1 = (1,1,1)$. Then $S_1 = S \cup \{(1,1,1)\} = \{(2,3,1), (1,1,1)\}$ is a linearly independent subset of $\mathbf{R}^3$ containing 2 vectors.

Now $[S_1] = \{\alpha(2,3,1) + \beta(1,1,1) \mid \alpha, \beta \in \mathbf{R}\}$

$$= \{(2\alpha + \beta, 3\alpha + \beta, \alpha + \beta) \mid \alpha, \beta \in \mathbf{R}\}$$

Now select $v_2 \in \mathbf{R}^3$ such that $v_2 \notin [S_1]$. We can take $v_2 = (3,4,0)$. How do we 'hit upon' this v_2? There are many ways. What we have done here is to take $\alpha = 1 = \beta$, then $2\alpha + \beta = 3$, $3\alpha + \beta = 4$, $\alpha + \beta = 2$. So $(3,4,2)$ belongs to $[S_1]$ Then, by changing the third component from 2 to 0, we get $(3,4,0)$, which is not in $[S_1]$. Since $v_2 \notin [S_1]$, $S_1 \cup \{v_2\}$ is linearly independent. That is, $S_2 = \{(2,3,1), (1,1,1), (3,4,0)\}$ is a linearly independent subset of $\mathbf{R}^3$. Since S_2 contains 3 vectors and $\dim_{\mathbf{R}} \mathbf{R}^3 = 3$, S_2 is a basis of $\mathbf{R}_3$.

Note : Since we had a large number of choices for both v_1 and v_2, it is obvious that we could have extended S to get a basis of $\mathbf{R}^3$ in many ways.

Example 12: For the vector space P_2 of all polynomials of degree ≤ 2, complete the linearly independent subset $S = \{x+1, 3x+2\}$ to form a basis of $\mathbf{P}_2$.

Solution: We note that P_2 has dimension 3, a basis being $\{1, x, x^2\}$ (see E19). So we have to add only one polynomial to S to get a basis of $\mathbf{P}_2$.

Now $[S] = \{a(x+1) + b(3x + 2) \mid a, b \in \mathbf{R}\}$

$$= \{(a + 3b)x + (a+2b) \mid a, b \in \mathbf{R}\}.$$

This shows that [S] does not contain any polynomial of degree 2. So we can choose $x^2 \in P_2$ because $x^2 \notin [S]$. So S can be extended to $\{x+1, 3x+2, x^2\}$, which is a basis of $\mathbf{P}_2$.

Have you wondered why there is no constant term in this basis? A constant term is not necessary. Observe that 1 is a linear combination of $x+1$ and $3x+2$, namely, $1 = 3(x+1) - 1(3x+2)$. So, $1 \in [S]$ and hence, $\forall \alpha \in \mathbf{R}$, $\alpha . 1 = \alpha \in [S]$.

E E20) Complete $S = \{(-3, 1/3)\}$ to a basis of $\mathbf{R}^2$.

E E21) Complete $S = \{(1,0,1), (2,3,-1)\}$ in two different ways to get two distinct bases of R^3.

E E22) For the vector space P_3, of all polynomials of degree ≤ 3, complete

a) $S = \{2, x^2 + x, 3x^3\}$

b) $S = \{x^2 + 2, x^2 - 3x^3\}$

to get a basis of P_3.

Let us now look at some properties of the dimensions of some subspaces.

4.5 DIMENSIONS OF SOME SUBSPACES

In Unit 3 you learnt what a subspace of a space is. Since it is a vector space itself, it must have a dimension. We have the following theorem.

Theorem 12: Let V be a vector space over a field F such that dim V = n. Let W be a subspace of V. Then dim W $\leq$ n.

Proof: Since W is a vector space over F in its own right, it has a basis. Suppose dim W = m. Then the number of elements in W's basis is m. These elements form a linearly independent subset of W, and hence, of V. Therefore, by Theorem 7, m $\leq$ n.

Remarks: If W is a subspace of V such that dim W = dim V = n, then W = V, since the basis of W is a set of linearly independent elements in V, and we can appeal to Theorem 5, Cor. 1.

Example 13: Let V be a subspace of R^2. What are the possible dimensions of V?

Solution: By Theorem 12, since dim $R^2 = 2$, the only possibilities for dim V are 0, 1 and 2.

If dim V = 2, then, by the remark above, $V = R^2$.
If dim V = 1, then $\{(\beta_1, \beta_2)\}$ is a basis of V, where $(\beta_1, \beta_2) \in R^2$. Then $V = \{\alpha(\beta_1, \beta_2) \mid \alpha \in R\}$.
This is a straight line that passes through the origin (since $0 \in V$).

If dim V = 0, then V = {0}.

Now try the following exercise.

E23) Let V be a subspace of $\mathbf{R}^3$. What are the 4 possibilities of its structure?

Now let us go further and discuss the dimension of the sum of subspaces (see Sec. 3.6).

If U and W are subspaces of a vector space V, then so are U+W and U∩W. Thus, all these subspaces have dimensions. We relate these dimensions in the following theorem.

Theorem 13: If U and W are two subspaces of a finite-dimensional vector space V over a field **F**, then
dim (U+W) = dim U + dim W − dim (U∩W).

Proof: We recall that $U + W = \{u + w \mid u \in U, w \in W\}$.

Let dim (U∩W) = r, dim U = m, dim W = n. We have to prove that dim (U+W) = m+n−r.

Let $\{v_1, v_2, \ldots\ldots, v_r\}$ be a basis of U∩W. Then $\{v_1, v_2, \ldots\ldots, v_r\}$ is a linearly independent subset of U and also of W. Hence, by Theorem 11, it can be extended to form a basis

$A = \{v_1, v_2, \ldots\ldots, v_r, u_{r+1}, u_{r+2}, \ldots\ldots, u_m\}$ of U and a basis
$B = \{v_1, v_2, \ldots\ldots, v_r, w_{r+1}, w_{r+2}, \ldots\ldots, w_n\}$ of W.

Now, note that none of the u's can be a w. For, if $u_s = w_t$ then $u_s \in U$, $w_t \in W$, so that $u_s \in U \cap W$. But then u_s must be a linear combination of the basis $\{v_1, \ldots\ldots, v_r\}$ of U∩W. This contradicts the fact that A is linearly independent. Thus,
$A \cup B = \{v_1, v_2, \ldots\ldots, v_r, u_{r+1}, \ldots\ldots, u_m, w_{r+1}, \ldots\ldots, w_n\}$, contains r+(m−r) + (n−r) vectors. We need to prove that A∪B is a basis of U+W. For this we first prove that A∪B is linearly independent, and then prove that every vector of U+W is a linear combination of A∪B. So let

$$\sum_{i=1}^{r} \alpha_i v_i + \sum_{j=r+1}^{m} \beta_j u_j + \sum_{k=r+1}^{n} \tau_k w_k = 0, \text{ where } \alpha_i, \beta_j, \tau_k \in F \ \forall\ i,j,k.$$

Then

$$\sum_{i=1}^{r} \alpha_i v_i + \sum_{j=r+1}^{m} \beta_j u_j = -\sum_{k=r+1}^{n} \tau_k w_k \qquad \ldots\ldots\ldots\ldots\ldots (1)$$

The vector on the left hand side of Equation (1) is a linear combination of $\{v_1, \ldots\ldots, v_r, u_{r+1}, \ldots\ldots, u_n\}$. So it is in U. The vector on the right hand side is in W. Hence, the vectors on both side of the equation are in U∩W. But $\{v_1, \ldots\ldots, v_r\}$ is a basis of U∩W. So the vectors on both sides of Equation (1) are a linear combination of the basis $\{v_1, \ldots\ldots, v_r\}$ of U∩W.

That is,

$$\sum_{i=1}^{r} \alpha_i v_i + \sum_{j=r+1}^{m} \beta_j u_j = \sum_{i=1}^{r} \delta_i v_i \qquad \ldots\ldots\ldots\ldots\ldots (2)$$

and

$$\sum_{k=r+1}^{n} \tau_k w_k = \sum_{i=1}^{r} \delta_i v_i \qquad \ldots\ldots\ldots\ldots (3)$$

where $\delta_i \in F \ \forall\ i = 1, \ldots\ldots, r$

(2) gives $\Sigma(\alpha_i - \delta_i)v_i + \Sigma \beta_j u_j = 0$.

But $\{v_1, \ldots, v_r, u_{r+1}, \ldots, u_m\}$ is linearly independent, so

$\alpha_i = \delta_i$ and $\beta_j = 0 \; \forall \; i,j$.

Similarly, since by (3)

$\Sigma \delta_i v_i + \Sigma \tau_k w_k = 0$,

we get $\delta_i = 0 \; \forall \; i$, $\tau_k = 0 \; \forall \; k$.

Since, we have already obtained $\alpha_i = \delta_i \; \forall \; i$, we get $\alpha_i = 0 \; \forall \; i$.

Thus, $\Sigma \alpha_i v_i + \Sigma \beta_j u_j + \Sigma \tau_k w_k = 0$

$\Rightarrow \alpha_i = 0, \beta_j = 0, \tau_k = 0 \; \forall \; i,j,k$.

So AUB is linearly independent.

Next, let $u + w \in U + W$.

Then $u = \Sigma \alpha_i v_i + \Sigma \beta_j u_j$

and $w = \Sigma \beta_i v_i + \Sigma \tau_k w_k$,

i.e., u+w is a linear combination of AUB.

$\therefore$, AUB is a basis of U+W, and

$\dim (U + W) = m + n - r = \dim U + \dim W - \dim (U \cap W)$

We give a corollary to Theorem 13 now.

Corollary: $\dim (U \oplus W) = \dim U + \dim W$.

Proof: The direct sum $U \oplus W$ indicates that $U \cap W = \{0\}$. Therefore, $\dim (U \cap W) = 0$. Hence, $\dim (U+W) = \dim U + \dim W$.

Let us use Theorem 13 now.

Example 14: Suppose U and W are subspaces of V, $\dim U = 4$, $\dim W = 5$, $\dim V = 7$. Find the possible values of $\dim (U \cap W)$.

Solution: Since W is a subspace of U+W, we must have $\dim (U+W) \geq \dim W = 5$. i.e., $\dim U + \dim W - \dim (U \cap W) \geq 5 \Rightarrow 4+5 - \dim (U \cap W) \geq 5 \Rightarrow \dim (U \cap W) \leq 4$.

On the other hand, U+W is a subspace of V, so $\dim (U+W) \leq 7$.

$\Rightarrow 5+4 - \dim (U \cap W) \leq 7$

$\Rightarrow \dim (U \cap W) \geq 2$

Thus, $\dim (U \cap W) = 2, 3$ or 4.

Example 15: Let V and W be the following subspaces of $\mathbf{R}^4$:

$V = \{(a,b,c,d) \mid b-2c+d = 0\}$, $W = \{(a,b,c,d) \mid a = d, b = 2c\}$

Find bases and the dimensions of V, W and $V \cap W$. Hence prove that $\mathbf{R}^4 = V + W$.

Solution: We observe that

$(a,b,c,d) \in V \Leftrightarrow b-2c+d = 0$.

$\Leftrightarrow (a,b,c,d) = (a,b,c,2c-b)$

$= (a,0,0,0) + (0,b,0,-b) + (0,0,c,2c)$

$= a\,(1,0,0,0) + b\,(0,1,0,-1) + c\,(0,0,1,2)$

This shows that every vector in V is a linear combination of the three linearly independent vectors (1,0,0,0), (0,1,0,−1), (0,0,1,2). Thus, a basis of V is

$A = \{(1,0,0,0), (0,1,0,-1), (0,0,1,2)\}$

Hence, $\dim V = 3$.

Next, $(a,b,c,d) \in W \Leftrightarrow a = d, b = 2c$

$\Leftrightarrow (a,b,c,d) = (a,2c,c,a) = (a,0,0,a) + (0,2c,c,0)$

$= a\,(1,0,0,1) + c\,(0,2,1,0)$,

which shows that W is generated by the linearly independent set $\{(1,0,0,1), (0,2,1,0)\}$.

$\therefore$, a basis for W is

$B = \{(1,0,0,1), (0,2,1,0)\}$,

and $\dim W = 2$.

Next, $(a,b,c,d) \in V \cap W \Leftrightarrow (a,b,c,d) \in V$ and $(a,b,c,d) \in W$

$\Leftrightarrow b - 2c + d = 0, a = d, b = 2c$

$\Leftrightarrow (a,b,c,d) = (0,2c,c,0) = c\,(0,2,1,0)$

Hence, a basis of $V \cap W$ is $\{(0,2,1,0)\}$ and $\dim(V \cap W) = 1$.

Finally, $\dim(V+W) = \dim V + \dim W - \dim(V \cap W)$
$= 3+2-1 = 4$.

Since $V + W$ is a subspace of R^4 and both have the same dimension,
$R^4 = V + W$.

E24) If U and W are 2-dimensional subspaces of R^3, show that $U \cap W \neq \{0\}$.

E25) If U and W are distinct 4-dimensional subspaces of a 6-dimensional vector space V, find the possible dimensions of $U \cap W$.

E26) Suppose V and W are subspaces of $\mathbf{R}^4$ such that $\dim V = 3$, $\dim W = 2$. Prove that $\dim(V \cap W) = 1$ or 2.

E27) Let V and W be subspaces of $\mathbf{R}^3$ defined as follows :
$V = \{(a,b,c) \mid b+2c = 0\}$
$W = \{(a,b,c) \mid a+b+c = 0\}$
a) Find bases and dimensions of V, W, $V \cap W$
b) Find $\dim(V+W)$.

Let us now look at the dimension of a quotient space. Before going further it may help to revise Sec. 3.7.

4.6 DIMENSION OF A QUOTIENT SPACE

In Unit 3 we defined the quotient space V/W for any vector space V and subspace W. Recall that $V/W = \{v + W \mid v \in V\}$.

We also showed that it is a vector space. Hence, it must have a basis and a dimension. The following theorem tells us what dim V/W should be.

Theorem 14: If W is a subspace of a finite-dimensional space V, then
$\dim (V/W) = \dim V - \dim W$.

Proof: Suppose dim V = n and dim W = m. Let $\{w_1, w_2, \ldots, w_m\}$ be a basis of W. Then there exist vectors $v_1, v_2, \ldots, v_k$ such that $\{w_1, w_2, \ldots, w_m, v_1, v_2, \ldots, v_k\}$ is a basis of V, where $m + k = n$ (Theorem 11).

We claim that $B = \{v_1 + W, v_2 + W, \ldots, v_k + W\}$ is a basis of V/W. First, let us show that B is linearly independent. For this, suppose

$\sum_{i=1}^{k} \alpha_i (v_i + W) = W$, where $\alpha_1, \ldots, \alpha_k$ are scalars

(note that the zero vector of V/W is W).

Then $\sum_{i=1}^{k} \alpha_i v_i + W = W$

$$\Rightarrow \left(\sum_{i=1}^{k} \alpha_i v_i\right) + W = W$$

$$\Rightarrow \sum_{i=1}^{k} \alpha_i v_i \in W$$

But $W = [\{w_1, w_2, \ldots, w_m\}]$, so

$$\sum_{i=1}^{k} \alpha_i v_i = \sum_{j=1}^{m} \beta_j w_j \text{ for some scalars } \beta_1, \ldots, \beta_m.$$

$$\Rightarrow \Sigma \alpha_i v_i - \Sigma \beta_j w_j = 0$$

But $\{w_1, \ldots, w_m, v_1, \ldots, v_k\}$ is a basis of V, so it is linearly independent. Hence we must have

$\beta_j = 0, \alpha_i = 0 \; \forall \; j, i.$

Thus,

$\Sigma \alpha_i (v_i + W) = W \Rightarrow \alpha_i = 0 \; \forall \; i.$

So B is linearly independent.

Next, to show that B generates V/W, let $v+W \in V/W$. Since $v \in V$ and $\{w_1, \ldots, w_m, v_1, \ldots, v_k\}$ is a basis of V,

$v = \sum_1^m \alpha_i w_i + \sum_1^k \beta_j v_j$, where the α_i s and β_j s are scalars.

Therefore,

$$v + W = \left(\sum_i \alpha_i w_i + \sum_j \beta_j v_j\right) + W$$

$$= \left\{\left(\sum_i \alpha_i w_i\right) + W\right\} + \left\{\left(\sum_j \beta_j v_j\right) + W\right\}$$

$$= W + \sum_1^k \beta_j (v_j + W), \text{ since } \sum \alpha_i w_i \in W.$$

$$= \sum_{j=1}^k \beta_j (v_j + W), \text{ since W is the zero element of V/W.}$$

Thus, $v + W$ is a linear combination of $\{v_j + W, j = 1, 2, \ldots, k\}$.

So, $v+W \in [B]$.

Thus, B is a basis of V/W.

Hence, $\dim V/W = k = n - m = \dim V - \dim W$.

Let us use this theorem to evaluate the dimensions of some familiar quotient spaces.

Example 16: If P_n denotes the vector space of all polynomials of degree $\leq n$, exhibit a basis of $\mathbf{P_4/P_2}$ and verify that $\dim \mathbf{P_4/P_2} = \dim \mathbf{P_4} - \dim \mathbf{P_2}$.

Solution: Now $\mathbf{P_4} = \{ax^4 + bx^3 + cx^2 + dx + e \mid a,b,c,d,e \in \mathbf{R}\}$ and
$\mathbf{P_2} = \{ax^2 + bx + c \mid a, b, c \in \mathbf{R}\}$.

Therefore, $\mathbf{P_4/P_2} = \{(ax^4 + bx^3) + \mathbf{P_2} \mid a, b \in \mathbf{R}\}$,

Now $(ax^4 + bx^3) + \mathbf{P_2}$

$= (ax^4 + \mathbf{P_2}) + (bx^3 + \mathbf{P_2})$

$= a(x^4 + \mathbf{P_2}) + b\,(x^3 + \mathbf{P_2})$

This shows that every element of $\mathbf{P_4/P_2}$ is a linear combination of the two elements $(x^4 + \mathbf{P_2})$ and $(x^3 + \mathbf{P_2})$.

These two elements of $\mathbf{P_4/P_2}$ are also linearly independent because if
$\alpha(x^4 + \mathbf{P_2}) + \beta\,(x^3 + \mathbf{P_2}) = \mathbf{P_2}$, then $\alpha x^4 + \beta x^3 \in \mathbf{P_2}$ $(\alpha, \beta \in \mathbf{R})$.

$\therefore$, $\alpha x^4 + \beta x^3 = ax^2 + bx + c$ for some $a,b,c \in \mathbf{R}$

$\Rightarrow \alpha = 0, \beta = 0, a = 0, b = 0, c = 0.$

Hence a basis of $\mathbf{P_4/P_2}$ is $\{x^4 + \mathbf{P_2}, x^3 + \mathbf{P_2}\}$.

Thus, $\dim (\mathbf{P_4/P_2}) = 2$. Also $\dim (\mathbf{P_4}) = 5$, $\dim (\mathbf{P_2}) = 3$, (see E19). Hence $\dim (\mathbf{P_4/P_2}) = \dim (\mathbf{P_4}) - \dim (\mathbf{P_2})$ is verified.

Try the following exercise now.

E28) Let V be an n-dimensional real vector space.
Find $\dim (V/V)$ and $\dim V/\{0\}$.

We end this unit by summarising what we have covered in it.

4.7 SUMMARY

In this unit, we have

1) introduced the important concept of linearly dependent and independent sets of vectors.
2) defined a basis of a vector space.
3) described how to obtain a basis of a vector space from a linearly dependent or a linearly independent subset of the vector space.
4) defined the dimension of a vector space.
5) obtained formulae for obtaining the dimension of the sum of two subspaces, intersection of two subspaces and quotient spaces.

4.8 SOLUTIONS/ANSWERS

E1) a) $a(1,2,3) + b(2,3,1) + c(3,1,2) = (0,0,0)$

$\Leftrightarrow (a,2a,3a) + (2b,3b,b) + (3c,c,2c) = (0,0,0)$

$\Leftrightarrow (a + 2b + 3c, 2a + 3b + c, 3a + b + 2c) = (0,0,0)$

$\Leftrightarrow$ $a + 2b + 3c = 0$ ....(1)

$2a + 3b + c = 0$ ....(2)

$3a + b + 2c = 0$ ....(3)

Then (1) + (2) − (3) gives $4b + 2c = 0$, i.e., $c = -2b$. Putting this value in (1) we get $a + 2b - 6b = 0$, i.e., $a = 4b$. Then (2) gives $8b + 3b - 2b = 0$, i.e., $b = 0$. Therefore, $a = b = c = 0$. Therefore, the given set is linearly independent.

b) $a(1,2,3) + b(2,3,1) + c(-3,-4,1) = (0,0,0)$

$\Leftrightarrow (a + 2b - 3c, 2a + 3b - 4c, 3a + b + c) = (0,0,0)$

$\Leftrightarrow a + 2b - 3c = 0$

$2a + 3b - 4c = 0$

$3a + b + c = 0.$

On simultaneously solving these equations you will find that a,b,c can have many non-zero values, one of them being $a = -1, b = 2, c = 1$. ∴ the given set is linearly dependent.

c) Linearly dependent.

d) Linearly independent.

E2) To show that {sin x, cos x} is linearly independent, suppose $a, b \in \mathbb{R}$ such that $a \sin x + b \cos x = 0$.

Putting $x = 0$ in this equation, we get $b = 0$. Now, take $x = \pi/2$.

We get $a = 0$. Therefore, the set is linearly independent.

Now, consider the equation

[illegible]

Since $\sin(x + \pi/6) = \sin x \cos \pi/6 + \cos x \sin \pi/6$

$= \sqrt{3}/2 \sin x + 1/2 \cos x$, taking $a = -\sqrt{3}/2$, $b = 1/2$, $c = 1$, we get a linear combination of the set $\{\sin x, \cos x, \sin(x + \pi/6)\}$ which shows that this set is linearly dependent.

E3) a) $ax^2 + b(x^2 + 1) = 0 \Rightarrow (a+b)x^2 + b = 0 \Rightarrow a + b = 0, b = 0$
$\Rightarrow a = 0, b = 0$. $\therefore$, the given set is linearly independent.

b) Linearly dependent because, for example,
$-5(x^2 + 1) + (x^2 + 11) + 2(2x^2 - 3) = 0$.

c) Linearly dependent.

d) Linearly dependent.

E4) Suppose $\alpha \in F$ such that $\alpha v = 0$. Then, from Unit 3 you know that $\alpha = 0$ or $v = 0$. But $v \neq 0$. $\therefore$, $\alpha = 0$, and $\{v\}$ is linearly independent.

E5) The set $S = \{(1,0), (0,1)\}$ is a linearly independent subset of $\mathbf{R}^2$. Now, suppose $\exists$ T such that $S \subseteq T \subseteq \mathbf{R}^2$. Let $(x,y) \in T$ such that $(x,y) \notin S$. Then we can always find $a,b,c \in \mathbf{R}$, not all zero, such that $a(1,0) + b(0,1) + c(x,y) = (0,0)$. (Take $a = -x$, $b = -y$, $c = 1$, for example.)

$\therefore S \cup \{(x,y)\}$ is linearly dependent. Since this is contained in T, T is linearly dependent

$\therefore$, the answer to the question in this exercise is 'No'.

E6) Let T be a finite subset of **P**. Suppose $1 \notin T$. Then, as in Example 4, $\exists$ non-zero $a_1, \ldots, a_k$ such that
$T = \{x^{a_1} + 1, \ldots, x^{a_k} + 1\}$.

Suppose $\sum_{i=1}^{k} \alpha_i (x^{a_i} + 1) = 0$, where $\alpha_i \in \mathbf{R}\ \forall\ i$.

Then $\alpha_1 x^{a_1} + \ldots + \alpha_k x^{a_k} + (\alpha_1 + \alpha_2 + \ldots + \alpha_k) = 0$
$\Rightarrow \alpha_1 = 0 = \alpha_2 = \ldots = \alpha_k$, so that T is linearly independent.
If $1 \in T$, then $T = \{1, x^{a_1} + 1, \ldots, x^{a_k} + 1\}$ for some non-zero $a_1, \ldots, a_k$.
Suppose,

$\beta_0 + \sum_{i=1}^{k} \beta_i (x^{a_i} + 1) = 0$, where $\beta_0, \beta_1, \ldots, \beta_k \in \mathbf{R}$.

Then $(\beta_0 + \beta_1 + \ldots + \beta_k) + \beta_1 x^{a_1} + \ldots + \beta_k x^{a_k} = 0$
$\Rightarrow \beta_0 + \beta_1 + \ldots + \beta_k = 0, \beta_1 = 0 = \beta_2 = \ldots = \beta_k$
$\Rightarrow \beta_0 = 0 = \beta_1 = \ldots = \beta_k$.
$\Rightarrow$ T is linearly independent.
Thus, every finite subset of $\{1, x+1, \ldots\}$ is linearly independent. Therefore, $\{1, x+1, \ldots\}$ is linearly independent.

E7) a) B is linearly independent and spans $\mathbf{R}^3$.

b) B is linearly independent.
For any $(a,b,c) \in \mathbf{R}^3$
we have $(a,b,c) = \frac{2b-a}{3}(1,2,0) + \frac{2a-b}{3}(2,1,0) + c(0,0,1)$.
Thus, B also spans $\mathbf{R}^3$.

E8) Firstly, any element of **P** is of the form $a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$, $a_i \in \mathbf{R}\ \forall\ i$. This is a linear combination of $\{1, x, \ldots, x^n\}$, a finite subset of the given set.
$\therefore$, the given set spans **P**. Secondly, Example 4 says that the given set is linearly independent.
$\therefore$ it is a basis of **P**.

E9) The set $\{1, x+1, x^2 + 2x\}$ is linearly independent. It also spans $\mathbf{P}_2$, since any element $a_0 + a_1 x + a_2 x^2 \in \mathbf{P}_2$ can be written as $(a_0 - a_1 + 2a_2) + (a_1 - 2a_2)(x+1) + a_2(x^2 + 2x)$. Thus, the set is a basis of $\mathbf{P}_2$.

E10) The set is linearly dependent, since $4-3(x+1)+(3x-1)+0 \cdot x^2=0$.
∴ it can't form a basis of P^2.

E11) a) We have to show that the given set is linearly independent.
Now $au+b(v+w)+c(w+t)+d(t+u)=0$, for $a,b,c,d \in F$.
$\Rightarrow (a+d)u+bv+(b+c)w+(c+d)t=0$
$\Rightarrow a+d=0$, $b=0$, $b+c=0$ and $c+d=0$, since $\{u,v,w,t\}$ is linearly independent. Thus, $a=0=b=c=d$.

∴ the given set is linearly independent. Since it has 4 vectors, it is a basis of V.

b) No, since $[\{u,t\}] \neq V$. For example, $w \notin [\{u,t\}]$ as $\{u,w,t\}$ is a linearly independent set by Theorem 2.

E12) You know that $\{1, x, x^2, x^3\}$ is a basis of P_3, and contains 4 vectors. The given set contains 6 vectors, and hence, by Theorem 7, it must be linearly dependent.

E13) A standard basis for R^3 is $\{(1,0,0), (0,1,0), (0,0,1)\}$. $\{1,x,x^2\}$ is a standard basis for P_2, because the coordinates of any vector $a_0+a_1x+a_2x^2$, in P^2, is (a_0, a_1, a_2).

E14) $\left(\frac{-13}{3}, \frac{-11}{3}, 2\right)$

E15) Since $0=0.v_1+0 \cdot v_2+\ldots+0 \cdot v_n$, the coordinates are $(0,0,\ldots,0)$.

E16) a) $6x+6=1.3+3(2x+1)+0 \cdot (x^2-2)$. ∴ the coordinates are (1,3,0)

b) (2/3, 1,1)

c) (2/3, 0, 1).

E17) Let $u=(a,b)$, $v=(c,d)$. We know that

$$(1,0)=\tfrac{1}{2}(a,b)+\tfrac{1}{2}(c,d)=\left(\frac{a+c}{2}, \frac{b+d}{2}\right),$$

and $(2,4)=3(a,b)-(c,d)=(3a-c, 3b-d)$.
∴ $a+c=2$, $b+d=0$, $3a-c=2$, $3b-d=4$. Solving these equations gives us $a=1$, $b=1$, $c=1$, $d=-1$
∴ $u=(1,2)$, $v=(1,-2)$.

E18) $C=\{x+iy \mid x,y \in R\}$. Consider the set $S=\{1+i0, 0+i1\}$. This spans C and is linearly independent. ∴ it is a basis of C. ∴ $\dim_R C=2$.

E19) The set $\{1,x,\ldots\ldots,x^n\}$ is a basis.

E20) We know that $\dim_R R^2=2$. ∴ we have to add one more vector to S to obtain a basis of R^2. Now $[S]=\{(-3\alpha, \frac{1}{3}\alpha) \mid \alpha \in R\}$.
∴ $(1,0) \notin [S]$. ∴ $\{(-3, 1/3), (1,0)\}$ is a basis of R^2.

E21) To obtain a basis we need to add one element. Now,
$[S]=\{\alpha(1,0,1)+\beta(2,3,-1) \mid \alpha,\beta \in R\}$
$=\{(\alpha+2\beta, 3\beta, \alpha-\beta) \mid \alpha, \beta \in R\}$
Then $(1,0,0) \notin [S]$ and $(0,1,0) \notin [S]$.

∴ $\{(1,0,1), (2,3,-1), (1,0,0)\}$ and $\{(1,0,1), (2,3,-1), (0,1,0)\}$ are two distinct bases of R^3.

E22) a) Check that $x \notin [S]$. ∴ $S \cup \{x\}$ is a basis.

b) $1 \notin [S]$. Let $S_1=S \cup \{1\}$. Then $x^3 \notin [S_1]$. Thus, a basis is $\{1, x^2+2, x^2-3x, x^3\}$.

E23) dim V can be 0,1,2 or 3. $\dim V=0 \Rightarrow V=\{0\}$.
$\dim V=1 \Rightarrow V=\{\alpha(\beta_1, \beta_2, \beta_3) \mid \alpha \in R\}$, for some $(\beta_1, \beta_2, \beta_3) \in R^3$.

This is a line in 3-dimensional space.
dim V = 2 ⟹ V is generated by two linearly independent space vectors. Thus, V is a plane.
dim V = 3 ⟹ V = $\mathbf{R}^3$.

E24) dim U = 2 = dim W. Now U+W is a subspace of $\mathbf{R}^3$.
∴ dim (U+W) ≤ 3, i.e., dim U + dim W − dim (U∩W) ≤ 3.
i.e., dim (U∩W) ≥ 1 ∴ U∩W ≠ {0}.

E25) dim V = 6, dim U = 4 = dim W and U ≠ W. Then dim (U+W) ≤ 6
⟹ 4+4 − dim (U∩W) ≤ 6 ⟹ dim (U∩W) ≥ 2. Also, dim (U+W) ≥ dim U
⟹ dim (U∩W) ≤ 4. ∴, the possible dimensions of U∩W are 2,3,4.

E26) Since V+W is a subspace of $\mathbf{R}^4$, dim (V+W) ≤ 4.
That is, dim V + dim W − dim (V∩W) ≤ 4.
∴, dim (V∩W) ≥ 1.
Also V∩W is a subspace of W. ∴, dim (V∩W) ≤ dim W = 2.
∴, 1 ≤ dim (V∩W) ≤ 2.

E27) a) Any element of V is v = (a,b,c) with b+2c = 0.
∴, v = (a,−2c,c) = a(1,0,0) + c(0,−2,1).
∴, a basis of V is {(1,0,0), (0,−2,1)}. ∴, dim V = 2.
Any element of W is w = (a,b,c) with a+b+c = 0.
∴, w = (a,b,−a−b) = a(1,0,−1) + b(0,1,−1)
∴, a basis of W is {(1,0,−1), (0,1,−1)}.
∴, dim W = 2.
Any element of V∩W is x = (a,b,c) with b+2c = 0 and a+b+c = 0.
∴, x = (a,−2c,c) with a − 2c + c = 0, that is, a = c.
∴, x = (c,−2c,c) = c(1,−2,1). ∴, a basis of V ∩ W is {1, − 2, 1}.
∴, dim (V ∩ W) = 1.

b) dim (V+W) = dim V + dim W − dim (V ∩ W) = 3. ∴, V + W = $\mathbf{R}^3$.

E28) 0, n.

Block

2

LINEAR TRANSFORMATIONS AND MATRICES

BLOCK 2 LINEAR TRANSFORMATIONS AND MATRICES

In the last block we introduced you to the basic algebraic structure that this course is built on, namely, vector spaces. In this block, consisting of 4 units, we will study certain functions between vector spaces, as well as matrices associated with these functions.

Functions, whose domain and co-domain are vector spaces, and have an added property of linearity, are called linear transformations, linear operators or linear mappings. In Units 5 and 6, we define and discuss some properties of such functions.

In Unit 7 we define a concept closely linked to linear transformations, namely, a matrix. We introduce you to various kinds of matrices, and show you how to obtain matrices associated with linear operators. We go on to introduce you to matrix multiplication, a concept that you may require to spend some time on, to properly digest it. This unit may take you a little longer to take in, but don't let that worry you. Do spend as much time as you need, because the ideas that we have dealt with in it are interesting and applicable in mathematics, as well as the other sciences.

In Unit 8 we define the rank of a matrix and relate it to the rank of its associated linear transformation. You will also learn the technique of reducing a matrix to echelon form for solving a system of linear equations.

At the end of the block we have given the media note accompanying our video programme 'Linear Transformations and Matrices'. Please do go through the note and view the programme at your study centre.

In the next block you will be needing a thorough knowledge of the material covered in this block. So, go through this block carefully.

A piece of advice before you get to grips with the units. Whenever a reference to an earlier section or result is made, it is best to pause, go back to the result referred to and carefully try to see the connection.

NOTATIONS AND SYMBOLS

$\dim_F V$	dimension of V over F
$\dim V$	dimension of V over F (F is understood here)
$R(T)$	range space of the operator T
$\operatorname{Ker} T$	kernel of the operator T
$L(U, V)$	set of all linear transformations from U to V
$A(V)$	$L(V, V)$
V^*	$L(V, F)$, where V is a vector space over F
$S \circ T$	composition of the mappings S and T
$\simeq$	is isomorphic to
δ_{ij}	Kronecker delta function
$F[x]$	set of all polynomials in x with coefficients in F
$\deg f$	degree of the polynomial f
$p(T)$	the operator obtained by substituting T for x in the polynomial $p(x)$
$M_{m \times n}(F)$	set of all m x n matrices over F
$M_n(F)$	set of all n x n matrices over F
$[T]_{B_1, B_2}$	matrix of T with respect to the bases B_1 and B_2
$[T]_B$	matrix of T with respect to the basis B
$\rho(A)$	rank of the matrix A
$\min(m,n)$	minimum of m and n
$[S]$	linear span of the set S
$R_{ij}(A)$ $(C_{ij}(A))$	elementary operation of interchanging the ith and jth rows (columns) of A
$R_i(a)(A)$ $(C_i(a)(A))$	elementary operation of multiplying the ith row (column) of A by a
$R_{ij}(a)(A)$ $(C_{ij}(a)(A))$	elementary operation of adding a times the jth row (column) to the ith row (column) of A
$A \overset{R}{\sim} B$	apply operation R on A to get B
A^θ, A^*, $\overline{A}^t$	Conjugate transpose of the square matrix A

UNIT 5 LINEAR TRANSFORMATIONS - I

Structure

5.1 INTRODUCTION

You have already learnt about a vector space and several concepts related to it. In this unit we initiate the study of certain mappings between two vector spaces, called linear transformations. The importance of these mappings can be realised from the fact that, in the calculus of several variables, every continuously differentiable function can be replaced, to a first approximation, by a linear one. This fact is a reflection of a general principle that every problem on the change of some quantity under the action of several factors can be regarded, to a first approximation, as a linear problem. It often turns out that this gives an adequate result. Also, in physics it is important to know how vectors behave under a change of the coordinate system. This requires a study of linear transformations.

In this unit we study linear transformations and their properties, as well as two spaces associated with a linear transformation, and their dimensions. Then, we prove the existence of linear transformations with some specific properties. We discuss the notion of an isomorphism between two vector spaces, which allows us to say that all finite-dimensional vector spaces of the same dimension are the 'same', in a certain sense.

Finally, we state and prove the Fundamental Theorem of Homomorphism and some of its corollaries, and apply them to various situations.

Since this unit uses concepts developed in Units 1, 3 and 4, we suggest that you revise these units before going further.

Objectives

After reading this unit, you should be able to

- verify the linearity of certain mappings between vector spaces;
- construct linear transformations with certain specified properties;
- calculate the rank and nullity of a linear operator;
- prove and apply the Rank Nullity Theorem;
- define an isomorphism between two vector spaces;
- show that two vector spaces are isomorphic if and only if they have the same dimension;
- prove and use the Fundamental Theorem of Homomorphism.

5.2 LINEAR TRANSFORMATIONS

In Unit 2 you came across the vector spaces R^2 and R^3. Now consider the mapping $f: R^2 \rightarrow R^3: f(x,y) = (x,y,0)$ (see Fig. 1).

f is a well defined function. Also notice that

i) $f((a,b) + (c,d)) = f((a + c, b + d)) = (a + c, b + d, 0) = (a,b,0) + (c,d,0)$

$= f((a,b)) + f((c,d))$, for (a,b), (c,d) in R, and

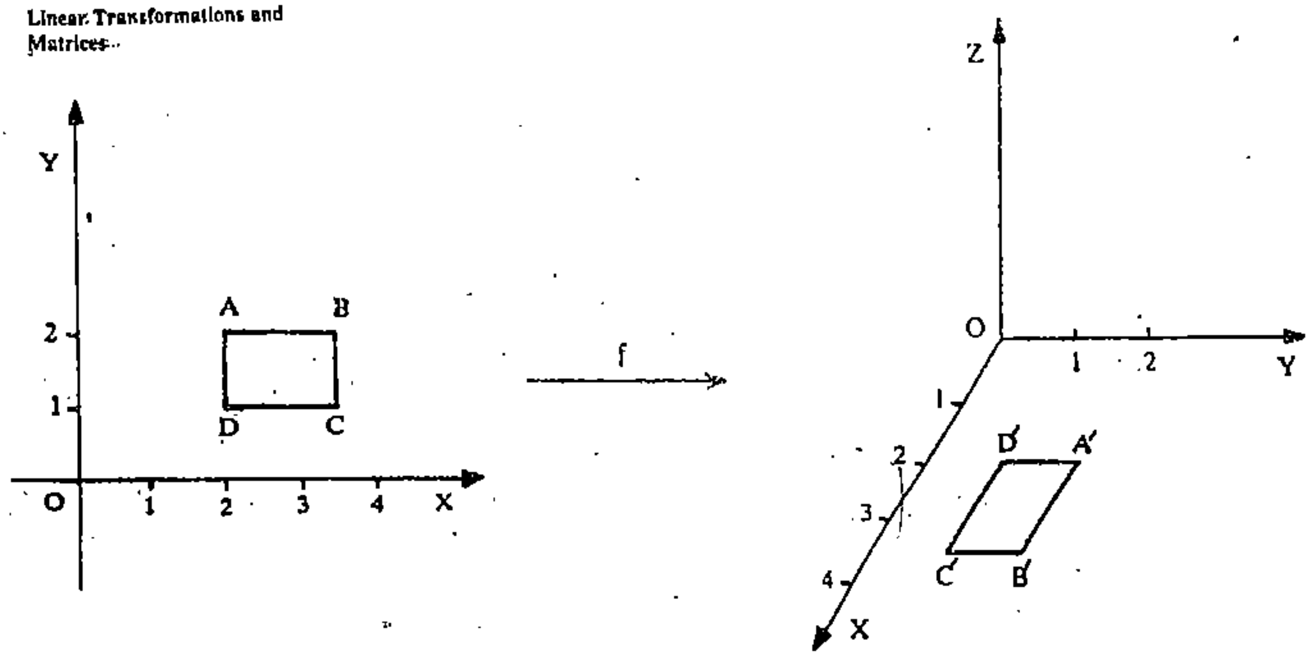


Fig. 1: f transforms ABCD to A'B'C'D'.

ii) for any $\alpha \in \mathbf{R}$ and $(a,b) \in \mathbf{R}^2$, $f(\alpha(a,b)) = f((\alpha a, \alpha b)) = (\alpha a, \alpha b, 0) = \alpha(a,b,0)$
$= \alpha f((a,b))$.

So we have a function f between two vector spaces such that (i) and (ii) above hold true.

(i) says that the sum of two plane vectors is mapped under f to the sum of their images under f. (ii) says that a line in the plane $\mathbf{R}^2$ is mapped under f to a line in $\mathbf{R}^3$.

The properties (i) and (ii) together say that f is linear, a term that we now define.

Definition: Let U and V be vector spaces over a field **F**. A **linear transformation** (or **linear operator**) from U to V is a function $T: U \to V$, such that

LT1) $T(u_1 + u_2) = T(u_1) + T(u_2)$, for $u_1, u_2 \in U$, and

LT2) $T(\alpha u) = \alpha T(u)$ for $\alpha \in \mathbf{F}$ and $u \in U$.

The conditions LT1 and LT2 can be combined to give the following equivalent condition.

LT3) $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in \mathbf{F}$ and $u_1, u_2 \in U$.

What we are saying is that [LT1 and LT2] $\Leftrightarrow$ LT3. This can be easily shown as follows:

We will show that LT3 $\Rightarrow$ LT1 and LT3 $\Rightarrow$ LT2. Now, LT3 is true $\forall \alpha_1, \alpha_2 \in \mathbf{F}$. Therefore, it is certainly true for $\alpha_1 = 1 = \alpha_2$, that is, LT1 holds.

Now, to show that LT2 is true, consider $T(\alpha u)$ for any $\alpha \in \mathbf{F}$ and $u \in U$. We have $T(\alpha u) = T(\alpha u + 0.u) = \alpha T(u) + 0.T(u) = \alpha T(u)$, thus proving that LT2 holds.

You can try and prove the converse now. That is what the following exercise is all about!

E E1) Show that the conditions LT1 and LT2 together imply LT3.

Before going further, let us note two properties of any linear transformation $T: U \to V$, which follow from LT1 (or LT2, or LT3).

LT4) $T(0) = 0$. Let's see why this is true. Since $T(0) = T(0 + 0) = T(0) + T(0)$ (by LT1), we subtract $T(0)$ from both sides to get $T(0) = 0$.

LT5) $T(-u) = -T(u) \forall u \in U$. Why is this so? Well, since $0 = T(0) = T(u - u)$
$= T(u) + T(-u)$, we get $T(-u) = -T(u)$.

E E 2) Can you show how LT4 and LT5 will follow from LT2?

Now let us look at some common linear transformations.

Example 1: Consider the vector space U over a field F, and the function $T : U \to U$ defined by $T(u) = u$ for all $u \in U$.

Show that T is a linear transformation. (This transformation is called the **identity transformation** and is denoted by I_U, or just I, if the *underlying* vector space is understood.)

Solution: For any $\alpha, \beta \in F$ and $u_1, u_2 \in U$, we have

$T(\alpha u_1 + \beta u_2) = \alpha u_1 + \beta u_2 = \alpha T(u_1) + \beta T(u_2)$.

Hence, LT3 holds, and T is a linear transformation.

Example 2: Let $T : U \to V$ be defined by $T(u) = 0$ for all $u \in U$.

Check that T is a linear transformation. (It is called the **null or zero transformation**, and is denoted by 0.)

Solution: For any $\alpha, \beta \in F$ and $u_1, u_2 \in U$, we have

$T(\alpha u_1 + \beta u_2) = 0 = \alpha.0 + \beta.0 = \alpha T(u_1) + \beta T(u_2)$.

Therefore, T is a linear transformation.

Example 3: Consider the function $pr_1 : \mathbf{R}^n \to \mathbf{R}$, defined by $pr_1[(x_1, \ldots\ldots, x_n)] = x_1$. Show that this is a linear transformation. (This is called the **projection on the first coordinate**. Similarly, we can define $pr_i : \mathbf{R}^n \to \mathbf{R}$ by $pr_i[(x_1, \ldots, x_{i-1}, x_i, \ldots, x_n)] = x_i$ to be the **projection on the i^{th} coordinate** for $i = 2, \ldots, n$. For instance, $pr_2 : \mathbf{R}^3 \to \mathbf{R} : pr_2(x,y,z) = y$.)

Solution : We will use LT3 to show that pr_1 is a linear operator. For $\alpha, \beta \in \mathbf{R}$ and $(x_1, \ldots, x_n), (y_1, \ldots, y_n)$ in $\mathbf{R}^n$, we have

$pr_1[(\alpha(x_1, \ldots\ldots, x_n) + \beta(y_1, \ldots\ldots, y_n)]$

$= pr_1(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \ldots\ldots, \alpha x_n + \beta y_n) = \alpha x_1 + \beta y_1$

$= \alpha pr_1[(x_1, \ldots, x_n)] + \beta pr_1[(y_1, \ldots\ldots, y_n)]$.

Thus pr_1 (and similarly pr_i) is a linear transformation.

Before going to the next example, we make a remark about projections.

Remark : Consider the function $p : \mathbf{R}^3 \to \mathbf{R}^2 ; p(x,y,z) = (x,y)$. This is a projection from $\mathbf{R}^3$ on to the xy-plane. Similarly, the functions f and g, from $\mathbf{R}^3 \to \mathbf{R}^2$, defined by $f(x,y,z) = (x,z)$ and $g(x,y,z) = (y,z)$ are projections from $\mathbf{R}^3$ onto the xz-plane and the yz-plane, respectively.

In general, any function $\theta : \mathbf{R}^n \to \mathbf{R}^m$ $(n > m)$, which is defined by dropping any $(n - m)$ coordinates, is a projection map.

Now let us see another example of a linear transformation that is very geometric in nature.

Example 4: Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be defined by $T(x,y) = (x,-y) \forall x,y \in \mathbf{R}$.
Show that T is a linear transformation.
(This is the **reflection** in the x-axis that we show in Fig. 2.)

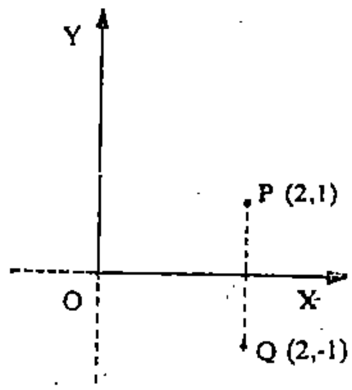


Fig. 2: Q is the reflection of P in the x-axis.

Solution: For $\alpha, \beta \in R$ and $(x_1,y_1), (x_2,y_2) \in R^2$, we have

$$T[\alpha(x_1,y_1) + \beta(x_2,y_2)] = T(\alpha x_1 + \beta x_2, \alpha y_1 + \beta y_2) = (\alpha x_1 + \beta x_2, -\alpha y_1 - \beta y_2)$$
$$= \alpha(x_1, -y_1) + \beta(x_2, -y_2)$$
$$= \alpha T(x_1,y_1) + \beta T(x_2,y_2).$$

Therefore, T is a linear transformation.

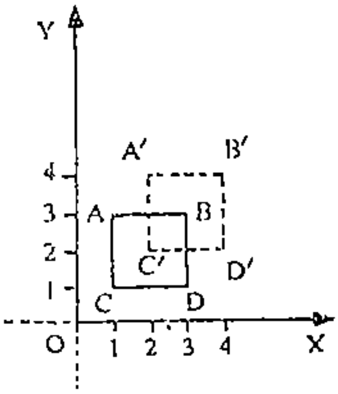


Fig. 3: A′B′C′D′ is the translation of ABCD by (1, 1).

So far we've given examples of linear transformations. Now we give an example of a very important function which is not linear. This example's importance lies in its geometric applications.

Example 5: Let u_0 be a fixed non-zero vector in U. Define T : U → U by $T(u) = u + u_0 \; \forall u \in U$. Show that T is not a linear transformation. (T is called the **translation** by u_0. See Fig. 3 for a geometrical view.)

Solution: T is not a linear transformation since LT4 does not hold. This is because $T(0) = u_0 \neq 0$.

Now, try the following exercises.

E E3) Let $T: R^2 \to R^2$ be the reflection in the y-axis. Find an expression for T as in Example 4. Is T a linear operator?

E E4) For a fixed vector (a_1, a_2, a_3) in R^3, define the mapping $T: R^3 \to R$ by $T(x_1, x_2, x_3) = a_1x_1 + a_2x_2 + a_3x_3$. Show that T is a linear transformation. Note that $T(x_1,x_2,x_3)$ is the dot product of (x_1,x_2,x_3) and (a_1, a_2, a_3) (ref. Sec. 2.4).

E E5) Show that the map $T : R^3 \to R^3$ defined by
$T(x_1, x_2, x_3) = (x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3)$ is a linear operator.

You came across the real vector space P_n, of all polynomials of degree less than or equal to n, in Unit 4. The next exercise concerns it.

E E6) Let $f \in P_n$ be given by

$f(x) = \alpha_0 + \alpha_1 x + \ldots\ldots + \alpha_n x^n, \alpha_i \in R \; \forall i.$

We define $(Df)(x) = \alpha_1 + 2\alpha_2 x + \ldots + n\alpha_n x^{n-1}$

Show that $D:P_n \rightarrow P_n$ is a linear transformation. (Observe that Df is nothing but the derivative of f. D is called the **differentiation operator**.)

In Unit 3 we introduced you to the concept of a quotient space. We now define a very useful linear transformation, using this concept.

Example 6: Let W be a subspace of a vector space U over a field F. W gives rise to the quotient space U/W. Consider the map $T:U \rightarrow U/W$ defined by $T(u) = u + W$.
T is called the **quotient map** or the **natural map**.
Show that T is a linear transformation.

Solution: For $\alpha, \beta \in F$ and $u_1, u_2 \in U$ we have

$$T(\alpha u_1 + \beta u_2) = (\alpha u_1 + \beta u_2) + W = (\alpha u_1 + W) + (\beta u_2 + W)$$
$$= \alpha(u_1 + W) + \beta(u_2 + W)$$
$$= \alpha T(u_1) + \beta T(u_2)$$

Thus, T is a linear transformation.

Now solve the following exercise, which is about plane vectors.

E7) Let $u_1 = (1,-1)$, $u_2 = (2,-1)$, $u_3 = (4,-3)$, $v_1 = (1,0)$, $v_2 = (0,1)$ and $v_3 = (1,1)$ be 6 vectors in R^2. Can you define a linear transformation $T: R^2 \rightarrow R^2$ such that $T(u_i) = v_i$, $i = 1, 2, 3$?

(Hint: Note that $2u_1 + u_2 = u_3$ and $v_1 + v_2 = v_3$.)

You have already seen that a linear transformation $T:U \rightarrow V$ must satisfy $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$. More generally, we can show that,

LT6 : $T(\alpha_1 u_1 + ... + \alpha_n u_n) = \alpha_1 T(u_1) + ... + \alpha_n T(u_n)$,

where $\alpha_i \in F$ and $u_i \in U$.

Let us show this by induction, that is, we assume the above relation for n = m, and prove it for m + 1. Now,

$$T(\alpha_1 u_1 + ... + \alpha_m u_m + \alpha_{m+1} u_{m+1})$$
$$= T(u + \alpha_{m+1} u_{m+1}), \text{ where } u = \alpha_1 u_1 + ... + \alpha_m u_m$$
$$= T(u) + \alpha_{m+1} T(u_{m+1}), \text{ since the result holds for } n = 2$$
$$= T(\alpha_1 u_1 + ... + \alpha_m u_m) + \alpha_{m+1} T(u_{m+1})$$
$$= \alpha_1 T(u_1) + ... + \alpha_m T(u_m) + \alpha_{m+1} T(u_{m+1}), \text{ since we have assumed the result for } n = m.$$

Thus, the result is true for n = m + 1. Hence, by induction, it holds true for all n.

Let us now come to a very important property of any linear transformation $T:U \rightarrow V$. In Unit 4 we mentioned that every vector space has a basis. Thus, U has a basis. We will now show that T is completely determined by its values on a basis of U. More precisely, we have

Theorem 1: Let S and T be two linear transformations from U to V, where $\dim_F U = n$. Let $\{e_1, ..., e_n\}$ be a basis of U. Suppose $S(e_i) = T(e_i)$ for i = 1, ..., n. Then $S(u) = T(u)$ for all $u \in U$.

Proof: Let $u \in U$. Since $\{e_1, ..., e_n\}$ is a basis of U, u can be uniquely written as $u = \alpha_1 e_1 + ... + \alpha_n e_n$, where the α_i are scalars.

Then, $S(u) = S(\alpha_1 e_1 + ... + \alpha_n e_n)$
$= \alpha_1 S(e_1) + ... + \alpha_n S(e_n)$, by LT6
$= \alpha_1 T(e_1) + ... + \alpha_n T(e_n)$
$= T(\alpha_1 e_1 + ... + \alpha_n e_n)$, by LT6
$= T(u)$.

What we have just proved is that once we know the values of T on a basis of U, then we can find T(u) for any $u \in U$.

Note: Theorem 1 is true even when U is not finite-dimensional. The proof, in this case, is on the same lines as above.

Let us see how the idea of Theorem 1 helps us to prove the following useful result.

Theorem 2: Let V be a real vector space and $T: \mathbf{R} \to V$ be a linear transformation. Then there exists $v \in V$ such that $T(\alpha) = \alpha v \ \forall \alpha \in \mathbf{R}$.

Proof: A basis for R is {1}. Let $T(1) = v \in V$. Then, for any $\alpha \in \mathbf{R}$, $T(\alpha) = \alpha T(1) = \alpha v$.

Once you have read Sec. 5.3 you will realise that this theorem says that $T(\mathbf{R})$ is a vector space of dimension one, whose basis is $\{T(1)\}$.

Now try the following exercise, for which you will need Theorem 1.

E E8) We define a linear operator $T: \mathbf{R}^2 \to \mathbf{R}^2$: $T(1,0) = (0,1)$ and $T(0,5) = (1,0)$. What is $T(3, 5)$? What is $T(5,3)$?

Now we shall prove a very useful theorem about linear transformations, which is linked to Theorem 1.

Theorem 3: Let $\{e_1, ..., e_n\}$ be a basis of U and let $v_1, ..., v_n$ be any n vectors in V. Then there exists one and only one linear transformation $T: U \to V$ such that $T(e_i) = v_i$, $i = 1, ..., n$.

Proof: Let $u \in U$. Then u can be uniquely written as $u = \alpha_1 e_1 + ... + \alpha_n e_n$ (see Unit 4, Theorem 9).

Define $T(u) = \alpha_1 v_1 + ... + \alpha_n v_n$. Then T defines a mapping from U to V such that $T(e_i) = v_i$ $\forall i = 1, ..., n$. Let us now show that T is linear. Let a, b be scalars and $u, u' \in U$. Then $\exists$ scalars $\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_n$ such that $u = \alpha_1 e_1 + ... + \alpha_n e_n$ and $u' = \beta_1 e_1 + ... + \beta_n e_n$.

Then $au + bu' = (a\alpha_1 + b\beta_1) e_1 + ... + (a\alpha_n + b\beta_n) e_n$.

Hence, $T(au + bu') = (a\alpha_1 + b\beta_1) v_1 + ... + (a\alpha_n + b\beta_n) v_n = a(\alpha_1 v_1 + ... + \alpha_n v_n) + b(\beta_1 v_1 + ... + \beta_n v_n) = aT(u) + bT(u')$.

Therefore, T is a linear transformation with the property that $T(e_i) = v_i \ \forall i$. Theorem 1 now implies that T is the only linear transformation with the above properties.

Let's see how Theorem 3 can be used.

Example 7: $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$ and $e_3 = (0, 0, 1)$ form the standard basis of $\mathbf{R}^3$. Let (1, 2), (2,3) and (3,4) be three vectors in $\mathbf{R}^2$. Obtain the linear transformation $T: \mathbf{R}^3 \to \mathbf{R}^2$ such that $T(e_1) = (1,2)$, $T(e_2) = (2,3)$ and $T(e_3) = (3,4)$.

Solution: By Theorem 3 we know that $\exists\, T: \mathbf{R}^3 \to \mathbf{R}^2$ such that $T(e_1) = (1,2)$, $T(e_2) = (2,3)$ and $T(e_3) = (3,4)$. We want to know what T(x) is, for any $x = (x_1, x_2, x_3) \in \mathbf{R}^3$. Now, $x = x_1 e_1 + x_2 e_2 + x_3 e_3$.

Hence, $T(x) = x_1 T(e_1) + x_2 T(e_2) + x_3 T(e_3)$
$= x_1(1,2) + x_2(2,3) + x_3(3,4)$
$= (x_1 + 2x_2 + 3x_3, 2x_1 + 3x_2 + 4x_3)$

Therefore, $T(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 2x_- \cdot 3x + 4x$ ... is the definition of the linear transformation T.

E9) Consider the complex field **C**. It is a vector space over R

a) What is its dimension over R? Give a basis of C over R.

b) Let $\alpha, \beta \in$ R. Give the linear transformation which maps the basis elements of C, obtained in (a), onto α and β, respectively.

Let us now look at some vector spaces that are related to a linear operator

5.3 SPACES ASSOCIATED WITH A LINEAR TRANSFORMATION

In Unit 1 you found that given any function, there is a set associated with it, namely, its range. We will now consider two sets which are associated with any linear transformation, T. These are the range and the kernel of T.

5.3.1 The Range Space and the Kernel

Let U and V be vector spaces over a field **F**. Let $T:U \to V$ be a linear transformation. We will define the range of T as well as the kernel of T. At first, you will see them as sets. We will prove that these sets are also vector spaces over F.

Definition: The **range of T**, denoted by **R(T)**, is the set $\{T(x) \mid x \in U\}$.
The **kernel** (or null space) of T, denoted by **Ker T**, is the set $\{x \in U \mid T(x) = 0\}$.
Note that $R(T) \subseteq V$ and $\text{Ker } T \subseteq U$.

To clarify these concepts consider the following examples.

Example 8: Let $I: V \to V$ be the identity transformation (see Example 1). Find R(I) and Ker I.

Solution: $R(I) = \{I(v) \mid v \in V\} = \{v \mid v \in V\} = V$. Also, $\text{Ker } I = \{v \in V \mid I(v) = 0\} = \{v \in V \mid v = 0\} = \{0\}$.

Example 9: Let $T:\mathbf{R}^3 \to \mathbf{R}$ be defined by $T(x_1, x_2, x_3) = 3x_1 + x_2 + 2x_3$. Find R(T) and Ker T.

Solution: $R(T) = \{x \in \mathbf{R} \mid \exists\, x_1, x_2, x_3 \in \mathbf{R} \text{ with } 3x_1 + x_2 + 2x_3 = x\}$.
For example, $0 \in R(T)$, since $0 = 3.0 + 0 + 2.0 = T(0,0,0)$

Also, $1 \in R(T)$, since $1 = 3.1/3 + 0 + 2.0 = T(1/3, 0, 0)$, or

$1 = 3.0 + 1 + 2.0 = T(0,1,0)$, or $1 = T(0,0,1/2)$, or $1 = T(1/6, 1/2, 0)$.

Now can you see that R(T) is the whole real line **R**? This is because, for any $\alpha \in$ **R**,

$\alpha = \alpha.1 = \alpha T(1/3,0,0) = T\left(\frac{\alpha}{3}, 0, 0\right) \in R(T)$.

$\text{Ker } T = \{(x_1, x_2, x_3) \in \mathbf{R}^3 \mid 3x_1 + x_2 + 2x_3 = 0\}$.

For example, $(0,0,0) \in$ Ker T. But $(1,0,0) \notin$ Ker T. $\therefore \text{Ker } T \neq \mathbf{R}^3$. In fact, Ker T is the plane $3x_1 + x_2 + 2x_3 = 0$ in $\mathbf{R}^3$.

Example 10: Let $T: \mathbf{R}^3 \to \mathbf{R}^3$ be defined by

$T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3)$.

Find R(T) and Ker T.

Solution: To find R(T), we must find conditions on $y_1, y_2, y_3 \in \mathbf{R}$ so that $(y_1, y_2, y_3) \in R(T)$, i.e., we must find some $(x_1, x_2, x_3) \in \mathbf{R}^3$ so that $(y_1, y_2, y_3) = T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 2x_1 + x_2, -x_1 - 2x_2 + 2x_3)$.

This means

$x_1 - x_2 + 2x_3 = y_1$ ....(1)

$2x_1 + x_2 = y_2$ ....(2)

$-x_1 - 2x_2 + 2x_3 = y_3$ ....(3)

Subtracting 2 times Equation (1) from Equation (2) and adding Equations (1) and (3) we get

$3x_2 - 4x_3 = y_2 - 2y_1$ ....(4)

and

$-3x_2 + 4x_3 = y_1 + y_3$ ....(5)

Adding Equations (4) and (5) we get

$y_2 - 2y_1 + y_1 + y_3 = 0$, that is, $y_2 + y_3 = y_1$.

Thus, $(y_1, y_2, y_3) \in R(T) \Rightarrow y_2 + y_3 = y_1$.

On the other hand, if $y_2 + y_3 = y_1$. We can choose

$$x_3 = 0,\ x_2 = \frac{y_2 - 2y_1}{3} \text{ and } x_1 = y_1 + \frac{y_2 - 2y_1}{3} = \frac{y_1 + y_2}{3}.$$

Then, we see that $T(x_1, x_2, x_3) = (y_1, y_2, y_3)$.

Thus, $y_2 + y_3 = y_1 \Rightarrow (y_1, y_2, y_3) \in R(T)$.

Hence, $R(T) = \{(y_1, y_2, y_3) \in \mathbf{R}^3 \mid y_2 + y_3 = y_1\}$

Now $(x_1, x_2, x_3) \in$ Ker T if and only if the following equations are true:

$x_1 - x_2 + 2x_3 = 0$.

$2x_1 + x_2 = 0$

$-x_1 - 2x_2 + 2x_3 = 0$

Of course $x_1 = 0, x_2 = 0, x_3 = 0$ is a solution. Are there other solutions? To answer this we proceed as in the first part of this example. We see that $3x_2 - 4x_3 = 0$. Hence, $x_3 = (3/4)\, x_2$.

Also, $2x_1 + x_2 = 0 \Rightarrow x_1 = -x_2/2$.

Thus, we can give arbitrary values to x_2 and calculate x_1 and x_3 in terms of x_2. Therefore, Ker T $= \{(-\alpha/2, \alpha, (3/4)\alpha) : \alpha \in \mathbf{R}\}$.

In this example, we see that finding R(T) and Ker T amounts to solving a system of equations. In Unit 9, you will learn a systematic way of solving a system of linear equations by the use of matrices and determinants

The following exercises will help you in getting used to R(T) and Ker T.

E E10) Let T be the zero transformation given in Example 2. Find Ker T and R(T). Does $1 \in R(T)$?

E E11) Find R(T) and Ker T for each of the following operators.

a) $T: \mathbf{R}^3 \to \mathbf{R}^2 : T(x, y, z) = (x, y)$

b) $T: \mathbf{R}^3 \to \mathbf{R} : T(x, y, z) = z$

c) $T: \mathbf{R}^3 \to \mathbf{R}^3 : T(x_1, x_2, x_3) = (x_1 + x_2 + x_3, x_1 + x_2 + x_3, x_1 + x_2 + x_3)$.

(Note that the operators in (a) and (b) are projections onto the xy-plane and the z-axis, respectively.)

Now that you are familiar with the sets R(T) and Ker T, we will prove that they are vector spaces.

Theorem 4: Let U and V be vector spaces over a field F. Let $T:U \to V$ be a linear transformation. Then Ker T is a subspace of U and R(T) is a subspace of V.

Proof: Let $x_1, x_2 \in$ Ker $T \subseteq U$ and $\alpha_1, \alpha_2 \in F$. Now, by definition, $T(x_1) = T(x_2) = 0$.
Therefore, $\alpha_1 T(x_1) + \alpha_2 T(x_2) = 0$

But $\alpha_1 T(x_1) + \alpha_2 T(x_2) = T(\alpha_1 x_1 + \alpha_2 x_2)$.
Hence, $T(\alpha_1 x_1 + \alpha_2 x_2) = 0$.
This means that $\alpha_1 x_1 + \alpha_2 x_2 \in$ Ker T.
Thus, by Theorem 4 of Unit 3, Ker T is a subspace of U.

Let $y_1, y_2 \in R(T) \subseteq V$, and $\alpha_1, \alpha_2 \in F$. Then, by definition of R(T), there exist $x_1, x_2 \in U$ such that $T(x_1) = y_1$ and $T(x_2) = y_2$.
So, $\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1 T(x_1) + \alpha_2 T(x_2)$
$= T(\alpha_1 x_1 + \alpha_2 x_2)$.
Therefore, $\alpha_1 y_1 + \alpha_2 y_2 \in R(T)$, which proves that R(T) is a subspace of V.

Now that we have proved that R(T) and Ker T are vector spaces, you know, from Unit 4, that they must have a dimension. We will study these dimensions now.

5.3.2 Rank and Nullity

Consider any linear transformation $T:U \to V$, assuming that dim U is finite. Then Ker T, being a subspace of U, has finite dimension and dim (Ker T) $\leq$ dim U. Also note that $R(T) = T(U)$, the image of U under T, a fact you will need to use in solving the following exercise.

E12) Let $\{e_1, \ldots, e_n\}$ be a basis of U. Show that R(T) is generated by $\{T(e_1), \ldots, T(e_n)\}$.

From E12 it is clear that, if dim U = n, then dim R(T) $\leq$ n.
Thus, dim R(T) is finite, and the following definition is meaningful.

Definition: The rank of T is defined to be the dimension of R(T), the range space of T. The nullity of T is defined to be the dimension of Ker T, the kernel (or the null space) of T.

Thus, rank (T) = dim R(T) and nullity (T) = dim Ker T.

We have already seen that rank (T) ≤ dim U and nullity (T) ≤ dim U.

Example 11: Let T:U → V be the zero transformation given in Example 2. What are the rank and nullity of T?

Solution: In E10 you saw that R(T) = {0} and Ker T = U. Therefore, rank (T) = 0 and nullity(T) = dim U.
Note that rank (T) + nullity (T) = dim U, in this case.

E13) If T is the identity operator on V, find rank (T) and nullity (T).

E14) Let D be the differentiation operator in E6. Give a basis for the range space of D and for Ker D. What are rank (D) and nullity (D)?

In the above example and exercises you will find that for T: U → V, rank (T) + nullity (T) = dim U. In fact, this is the most important result about rank and nullity of a linear operator. We will now state and prove this result.

This theorem is called the Rank Nullity Theorem.

Theorem 5: Let U and V be vector spaces over a field **F** and dim U = n. Let T:U → V be a linear operator. Then rank (T) + nullity (T) = n.

Proof: Let nullity (T) = m, that is, dim Ker T = m. Let $\{e_1,....,e_m\}$ be a basis of Ker T. We know that Ker T is a subspace of U. Thus, by Theorem 11 of Unit 4, we can extend this basis to obtain a basis $\{e_1,.....,e_m, e_{m+1},....., e_n\}$ of U. We shall show that $\{T(e_{m+1}), ..., T(e_n)\}$ is a basis of R(T). Then, our result will follow because dim R(T) will be
n − m = n − nullity (T).

Let us first prove that $\{T(e_{m+1}), ..., T(e_n)\}$ spans, or generates, R(T). Let y ∈ R(T). Then, by definition of R(T), there exists x ∈ U such that T(x) = y.

Let $x = c_1e_1 + ... + c_me_m + c_{m+1}e_{m+1} + ... + c_ne_n, c_i \in F\ \forall i$.

Then

$y = T(x) = c_1T(e_1) + ... + c_mT(e_m) + c_{m+1}T(e_{m+1}) + ... + c_nT(e_n)$

$= c_{m+1}T(e_{m+1}) + ... + c_nT(e_n)$,

because $T(e_1) = .. = T(e_m) = 0$, since $T(e_i) \in$ Ker T $\forall i = 1, ..., m$. ∴ any y ∈ R(T) is a linear combination of $\{T(e_{m+1}), ..., T(e_n)\}$. Hence, R(T) is spanned by $\{T(e_{m+1}), ..., T(e_n)\}$.
It remains to show that the set $\{T(e_{m+1}), ... T(e_n)\}$ is linearly independent. For this, suppose there exist $a_{m+1}, ..., a_n \in F$ with $a_{m+1}T(e_{m+1}) + ... + a_nT(e_n) = 0$.

Then, $T(a_{m+1}e_{m+1} + ... + a_ne_n) = 0$.

Hence, $a_{m+1}e_{m+1} + ... + a_ne_n \in$ Ker T, which is generated by $\{e_1, ..., e_m\}$.

Therefore, there exist $a_1, ..., a_m \in F$ such that

$a_{m+1}e_{m+1} + ... + a_ne_n = a_1e_1 + ... + a_me_m$.

$\Rightarrow (-a_1)e_1 + ... + (-a_m)e_m + a_{m+1}e_{m+1} + ... + a_ne_n = 0.$

Since $\{e_1, ..., e_n\}$ is a basis of U, it follows that this set is linearly independent. Hence,

$-a_1 = 0, \ldots, -a_m = 0, a_{m+1} = 0, \ldots, a_n = 0$. In particular, $a_{m+1} = \ldots = a_n = 0$, which we wanted to prove.

Therefore, dim R(T) = n − m = n − nullity (T), that is, rank (T) + nullity (T) = n.

Let us see how this theorem can be useful.

Example 12: Let $L: \mathbb{R}^3 \to \mathbb{R}$ be the map given by L(x, y, z) = x + y + z. What is nullity (L)?

Solution: In this case it is easier to obtain R(L), rather than Ker L. Since L (1, 0, 0) = 1 ≠ 0, R(L) ≠ {0}, and hence dim R(L) ≠ 0. Also, R(L) is a subspace of **R**. Thus, dim R(L) ≤ dim **R** = 1. Therefore, the only possibility for dim R(L) is dim R(L) = 1. By Theorem 5, dim Ker L + dim R(L) = 3.

Hence, dim Ker L = 3−1 = 2. That is, nullity (L) = 2.

E15) Give the rank and nullity of each of the linear transformations in E11.

E16) Let U and V be real vector spaces and T:U → V be a linear transformation, where dim U = 1. Show that R(T) is either a point or a line.

Before ending this section we will prove a result that links the rank (or nullity) of the composite of two linear operators with the rank (or nullity) of each of them.

Theorem 6: Let V be a vector space over a field F. Let S and T be linear operators from V to V. Then

a) rank (ST) ≤ min (rank (S), rank (T))
b) nullity (ST) ≥ max (nullity (S), nullity (T))

Proof: We shall prove (a). Note that (ST) (v) = S(T(v)) for any v ∈ V (you'll study more about compositions in Unit 6).

Now, for any y ∈ R (ST), ∃ v ∈ V such that,

y = (ST) (v) = S (T(v)) ....(1)

Now, (1) ⇒ y ∈ R(S).

Therefore, R(ST) ⊆ R(S). This implies that rank (ST) ≤ rank (S).

Again, (1) ⇒ y ∈ S (R(T)), since T(v) ∈ R(T).

∴ R (ST) ⊆ S (R (T)), so that

dim R (ST) ≤ dim S (R (T)) ≤ dim R (T) (since dim L (U) ≤ U, for any linear operator L).

Therefore, rank (ST) ≤ rank (T).

Thus, rank (ST) ≤ min (rank (S), rank (T)).

The proof of this theorem will be complete, once you solve the following exercise.

E 17) Prove (b) of Theorem 6 using the Rank Nullity Theorem.

We would now like to discuss some linear operators that have special properties.

5.4 SOME TYPES OF LINEAR TRANSFORMATIONS

Let us recall, from Unit 1, that there can be different types of functions, some of which are one-one, onto or invertible. We can also define such types of linear transformations as follows.

Definition: Let $T : U \to V$ be a linear transformation.

a) T is called one-one (or **injective**) if, for $u_1, u_2 \in U$ with $u_1 \neq u_2$, we have $T(u_1) \neq T(u_2)$. If T is injective, we also say T is 1 – 1.
   Note that **T is 1–1 if $T(u_1) = T(u_2) \Rightarrow u_1 = u_2$.**
b) T is called **onto** (or **surjective**) if, for each $v \in V$, $\exists\, u \in U$ such that $T(u) = v$, that is, **$R(T) = V$.**

Can you think of examples of such functions?
The identity operator is both one-one and onto. Why is this so? Well, $I: V \to V$ is an operator such that, if $v_1, v_2 \in V$ with $v_1 \neq v_2$ then $I(v_1) \neq I(v_2)$. Also, $R(I) = V$, so that I is onto.

E 18) Show that the zero operator $0: R \to R$ is not one-one.

An important result that characterises injectivity is the following:

Theorem 7: $T : U \to V$ is one-one if and only if $\text{Ker } T = \{0\}$.

Proof: First assume T is one-one. Let $u \in \text{Ker } T$. Then $T(u) = 0 = T(0)$. This means that $u = 0$. Thus, $\text{Ker } T = \{0\}$. Conversely, let $\text{Ker } T = \{0\}$. Suppose $u_1, u_2 \in U$ with $T(u_1) = T(u_2) \Rightarrow T(u_1 - u_2) = 0 \Rightarrow u_1 - u_2 \in \text{Ker } T \Rightarrow u_1 - u_2 = 0 \Rightarrow u_1 = u_2$. Therefore, T is 1 – 1.

Suppose now that T is a one-one and onto linear transformation from a vector space U to a vector space V. Then, from Unit 1 (Theorem 4), we know that T^{-1} exists.

But is T^{-1} linear? The answer to this question is 'yes', as is shown in the following theorem.

Theorem 8: Let U and V be vector spaces over a field F. Let $T : U \to V$ be a one-one and onto linear transformation. Then $T^{-1}: V \to U$ is a linear transformation.

In fact, T^{-1} is also 1–1 and onto.

$T^{-1}(y) = x \Leftrightarrow T(x) = y$

Proof: Let $y_1, y_2 \in V$ and $\alpha_1, \alpha_2 \in F$. Suppose $T^{-1}(y_1) = x_1$ and $T^{-1}(y_2) = x_2$. Then, by definition, $y_1 = T(x_1)$ and $y_2 = T(x_2)$.
Now, $\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1 T(x_1) + \alpha_2 T(x_2) = T(\alpha_1 x_1 + \alpha_2 x_2)$
Hence, $T^{-1}(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 x_1 + \alpha_2 x_2$
$= \alpha_1 T^{-1}(y_1) + \alpha_2 T^{-1}(y_2)$

This shows that T^{-1} is a linear transformation.

We will now show that T^{-1} is 1 – 1. For this, suppose $y_1, y_2 \in V$ such that $T^{-1}(y_1) = T^{-1}(y_2)$. Let $x_1 = T^{-1}(y_1)$ and $x_2 = T^{-1}(y_2)$.

Then $T(x_1) = y_1$ and $T(x_2) = y_2$. We know that $x_1 = x_2$. Therefore, $T(x_1) = T(x_2)$, that is, $y_1 = y_2$. Thus, we have shown that $T^{-1}(y_1) = T^{-1}(y_2) \Rightarrow y_1 = y_2$, proving that T^{-1} is 1 – 1. T^{-1} is also surjective because, for any $u \in U$, $\exists\ T(u) = v \in V$ such that $T^{-1}(v) = u$.

Theorem 8 says that a one-one and onto linear transformation is **invertible**, and the inverse is also a one-one and onto linear transformation.

This theorem immediately leads us to the following definition.

Definition: Let U and V be vector spaces over a field F, and let $T: U \to V$ be a one-one and onto linear transformation. Then T is called an **isomorphism** between U and V.

In this case we say that U and V are **isomorphic** vector spaces. This is denoted by $U \simeq V$.

An obvious example of an isomorphism is the identity operator. Can you think of any other? The following exercise may help.

E19) Let $T: \mathbf{R}^3 \to \mathbf{R}^3 : T(x, y, z) = (x + y, y, z)$. Is T an isomorphism? Why? Define T^{-1}, if it exists.

E20) Let $T: \mathbf{R}^3 \to \mathbf{R}^2 : T(x, y, z) = (x + y, y + z)$. Is T an isomorphism?

In all these exercises and examples, have you noticed that if T is an isomorphism between U and V then T^{-1} is an isomorphism between V and U?

Using these properties of an isomorphism we can get some useful results, like the following.

Theorem 9: Let $T: U \to V$ be an isomorphism. Suppose $\{e_1, \ldots, e_n\}$ is a basis of U. Then $\{T(e_1), \ldots, T(e_n)\}$ is a basis of V.

Proof: First we show that the set $\{T(e_1), \ldots, T(e_n)\}$ spans V. Since T is onto, $R(T) = V$. Thus, from E12 you know that $\{T(e_1), \ldots, T(e_n)\}$ spans V.

Let us now show that $\{T(e_1), \ldots, T(e_n)\}$ is linearly independent. Suppose there exist scalars $c_1, \ldots, c_n$, such that $c_1 T(e_1) + \ldots + c_n T(e_n) = 0$. .... (1)

We must show that $c_1 = \ldots = c_n = 0$.

Now, (1) implies that

$T(c_1 e_1 + \ldots + c_n e_n) = 0$.

Since T is one-one and $T(0) = 0$, we conclude that

$$c_1 e_1 + \ldots + c_n e_n = 0.$$

But $\{e_1, \ldots, e_n\}$ is linearly independent. Therefore,

$$c_1 = \ldots = c_n = 0.$$

Thus, we have shown that $\{T(e_1), \ldots, T(e_n)\}$ is a basis of V.

Remark: The argument showing the linear independence of $\{T(e_1), \ldots, T(e_n)\}$ in the abov theorem can be used to prove that any one-one linear transformation $T: U \to V$ maps any linearly independent subset of U onto a linearly independent subset of V (see E22).

We now give an important result equating 'isomorphism' with '1 – 1' and with 'onto' in the finite-dimensional case.

Theorem 10: Let $T : U \to V$ be a linear transformation where U, V are of the same finite dimension. Then the following statements are equivalent.

a) T is 1 – 1.

b) T is onto.

c) T is an isomorphism.

Proof: To prove the result we will prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). Let dim U = dim V = n.

Now (a) implies that Ker T = {0} (from Theorem 7). Hence, nullity (T) = 0. Therefore, by Theorem 5, rank (T) = n, that is, dim R (T) = n = dim V. But R (T) is a subspace of V. Thus, by the remark following Theorem 12 of Unit 4, we get R (T) = V, i.e., T is onto, i.e., (b) is true. So (a) $\Rightarrow$ (b).

Similarly, if (b) holds then rank (T) = n, and hence, nullity (T) = 0. Consequently, Ker T = {0}, and T is one-one. Hence, T is one-one and onto, i.e., T is an isomorphism. Therefore, (b) implies (c).

That (a) follows from (c) is immediate from the definition of an isomorphism.

Hence, our result is proved.

Caution: Theorem 10 is true for finite-dimensional spaces U and V, of the same dimension. It is not true, otherwise. Consider the following counter-example.

Example 13 (To show that the spaces have to be finite-dimensional): Let V be the real vector space of all polynomials. Let $D: V \to V$ be defined by $D(a_0 + a_1 x + \ldots + a_r x^r) = a_1 + 2a_2 x + \ldots + ra_r x^{r-1}$. Then show that D is onto but not 1 – 1.

Solution: Note that V has infinite dimension, a basis being $\{1, x, x^2, \ldots\}$. D is onto because any element of V is of the form $a_0 + a_1 x + \ldots + a_n x^n = D\left(a_0 x + \frac{a_1}{2}x^2 + \cdots + \frac{a_n}{n+1}x^{n+1}\right)$

D is not 1 – 1 because, for example, $1 \neq 0$ but D (1) = D (0) = 0.

The following exercise shows that the statement of Theorem 10 is false if dim U $\neq$ dim V.

E E21) Define a linear operator $T : \mathbf{R}^3 \to \mathbf{R}^2$ such that T is onto but T is not 1 – 1. Note that dim $\mathbf{R}^3 \neq$ dim $\mathbf{R}^2$.

Let us use Theorems 9 and 10 to prove our next result.

Theorem 11: Let $T : V \to V$ be a linear transformation and let $\{e_1, \ldots, e_n\}$ be a basis of V. Then T is one-one and onto if and only if $\{T(e_1), \ldots, T(e_n)\}$ is linearly independent.

Proof: Suppose T is one-one and onto. Then T is an isomorphism. Hence, by Theorem 9, $\{T(e_1), \ldots, T(e_n)\}$ is a basis. Therefore, $\{T(e_1), \ldots, T(e_n)\}$ is linearly independent.

Conversely, suppose $\{T(e_1), \ldots, T(e_n)\}$ is linearly independent. Since $\{e_1, \ldots, e_n\}$ is a basis of V, dim V = n. Therefore, any linearly independent subset of n vectors is a basis of V (by Unit 4, Theorem 5, Cor. 1). Hence, $\{T(e_1), \ldots, T(e_n)\}$ is a basis of V. Then, any element v of V is of the form $v = \sum_{i=1}^{n} c_i T(e_i) = T\left(\sum_{i=1}^{n} c_i e_i\right)$, where $c_1, \ldots, c_n$ are scalars. Thus, T is onto, and we can use Theorem 10 to say that T is an isomorphism.

Here are some exercises now.

E E 22) a) Let $T: U \to V$ be a one-one linear transformation and let $\{u_1, \ldots, u_k\}$ be a linearly independent subset of U. Show that the set $\{T(u_1), \ldots, T(u_k)\}$ is linearly independent.

b) Is it true that every linear transformation maps every linearly independent set of vectors into a linearly independent set?

c) Show that every linear transformation maps a linearly dependent set of vectors onto a linearly dependent set of vectors.

E23) Let $T: R^3 \to R^3$ be defined by $T(x_1, x_2, x_3) = (x_1 + x_3, x_2 + x_3, x_1 + x_2)$. Is T invertible? If yes, find a rule for T^{-1} like the one which defines T.

We have seen, in Theorem 9, that if $T: U \to V$ is an isomorphism, then T maps a basis of U onto a basis of V. Therefore, dim U = dim V. In other words, if U and V are isomorphic then dim U = dim V. The natural question arises whether the converse is also true. That is, if dim U = dim V, both being finite, can we say that U and V are isomorphic? The following theorem shows that this is indeed the case.

Theorem 12: Let U and V be finite-dimensional vector spaces over F. Then U and V are isomorphic if and only if dim U = dim V.

Proof: We have already seen that if U and V are isomorphic then dim U = dim V. Conversely, suppose dim U = dim V = n. We shall show that U and V are isomorphic. Let $\{e_1,...,e_n\}$ be a basis of U and $\{f_1,...,f_n\}$ be a basis of V. By Theorem 3, there exists a linear transformation $T: U \to V$ such that $T(e_i) = f_i$, i = 1, ..., n.

We shall show that T is 1 – 1.

Let $u = c_1 e_1 + ... + c_n e_n$ be such that $T(u) = 0$.

Then $0 = T(u) = c_1 T(e_1) + ... + c_n T(e_n)$

$= c_1 f_1 + ... + c_n f_n$.

Since $\{f_1, ..., f_n\}$ is a basis of V, we conclude that $c_1 = c_2 = ... = c_n = 0$. Hence, u = 0. Thus, Ker T = {0} and, by Theorem 7, T is one-one.

Therefore, by Theorem 10, T is an isomorphism, and $U \simeq V$.

An immediate consequence of this theorem follows.

Corollary: Let V be a real (or complex) vector space of dimension n. Then V is isomorphic to R^n (or C^n), respectively.

Proof: Since $\dim R^n = n = \dim_R V$, we get $V \simeq R^n$. Similarly, if $\dim_C V = n$, then $V \simeq C^n$.

We generalise this corollary in the following remark.

Remark: Let V be a vector space over F and let $B = \{e_1, ..., e_n\}$ be a basis of V. Each $v \in V$ can be uniquely expressed as $v = \sum_{i=1}^{n} \alpha_i e_i$. Recall that $\alpha_1, ...\alpha_n$ are called the coordinates of v with respect to B (refer to Sec. 4.4.1).

Define $\theta : V \rightarrow F^n : \theta(v) = (\alpha_1, ..., \alpha_n)$. Then θ is an isomorphism from V to F^n. This is because θ is 1 – 1, since the coordinates of v with respect to B are uniquely detemined. Thus, $V \simeq F^n$.

We end this section with an exercise.

E E 24) Let $T : U \rightarrow V$ be a one-one linear mapping. Show that T is onto if and only if dim U = dim V. (Of course, you must assume that U and V are finite-dimensional spaces.)

Now let us look at isomorphisms between quotient spaces.

5.5 HOMOMORPHISM THEOREMS

Linear transformations are also called vector space homomorphisms. There is a basic theorem which uses the properties of homomorphisms to show the isomorphism of certain quotient spaces (ref. Unit 3). It is simple to prove, but is very important because it is always being used to prove more advanced theorems on vector spaces. (In the Abstract Algebra course we will prove this theorem in the setting of groups and rings.)

This theorem is called the Fundamental Theorem of Homomorphism.

Theorem 13: Let V and W be vector spaces over a field F and $T: V \rightarrow W$ be a linear transformation. Then **V/Ker T $\simeq$ R (T).**

Proof: You know that Ker T is a subspace of V, so that V/Ker T is a well defined vector space over F. Also $R(T) = \{T(v) \mid v \in V\}$. To prove the theorem let us define θ**:V/ Ker T** $\rightarrow R(T)$ by $\theta(v + \text{Ker } T) = T(v)$.

Firstly, we must show that θ is a well defined function, that is, if v + Ker T = v' + Ker T then $\theta(v + \text{Ker } T) = \theta(v' + \text{Ker } T)$, i.e., T (v) = T (v').

Now, v + Ker T = v' + Ker T $\Rightarrow (v - v') \in$ Ker T (see Unit 3, E23)

$\Rightarrow T(v - v') = 0 \Rightarrow T(v) = T(v')$, and hence, θ is well defined.

Next, we check that θ is a linear transformation. For this, let $a, b \in F$ and $v, v' \in V$. Then $\theta\{a(v + \text{Ker } T) + b(v' + \text{Ker } T)\}$

$= \theta(av + bv' + \text{Ker } T)$ (ref. Unit 3)

$= T(av + bv')$

$= aT(v) + bT(v')$, since T is linear.

$= a\theta(v + \text{Ker } T) + b\theta(v' + \text{Ker } T)$.

Thus, θ is a linear transformation.

We end the proof by showing that θ is an isomorphism. θ is 1 – 1 (because $\theta(v + \text{Ker } T) = 0 \Rightarrow T(v) = 0 \Rightarrow v \in \text{Ker } T \Rightarrow v + \text{Ker } T = 0$ (in V/Ker T)

Thus, Ker $\theta = \{0\}$

θ is onto (because any element of R (T) is T (v) = θ (v + Ker T)).

So we have proved that θ is an isomorphism. This proves that V/Ker T $\simeq$ R (T).

Let us consider an immediate useful application of Theorem 13.

Example 14: Let V be a finite-dimensional space and let S and T be linear transformations from V to V. Show that

rank (ST) = rank (T) – dim (R (T) $\cap$ Ker S).

Solution: We have $V \xrightarrow{T} V \xrightarrow{S} V$. ST is the composition of the operators S and T, which you have studied in Unit 1, and will also study in Unit 6. Now, we apply Theorem 13 to the homomorphism $\theta : T(V) \to ST(V) : \theta(T(v)) = (ST)(v)$.

Now, Ker $\theta = \{x \in T(V) \mid S(x) = 0\}$ = Ker S $\cap$ T (V) = Ker S $\cap$ R (T)

Also R(θ) = ST (V), since any element of ST (V) is (ST) (v) = θ (T (v)). Thus,

$$\frac{T(V)}{\text{Ker } S \cap T(V)} \simeq ST(V)$$

Therefore,

$$\dim \frac{T(V)}{\text{Ker } S \cap T(V)} = \dim ST(V)$$

That is, dim T (V) – dim (Ker S $\cap$ T (V)) = dim ST (V), which is what we had to show.

E25) Using Example 14 and the Rank Nullity Theorem, show that
nullity (ST) = nullity (T) + dim (R (T) $\cap$ Ker S).

Now let us see another application of Theorem 13.

Example 15: Show that $R^3/R \simeq R^2$.

Solution : Note that we can consider R as a subspace of R^3 for the following reason: any element α of R is equated with the element (α, 0, 0) of R^3. Now, we define a function $f: R^3 \to R^2 : f(\alpha, \beta, \gamma) = (\beta, \gamma)$. Then f is a linear transformation and Ker f = $\{(\alpha, 0, 0) \mid \alpha \in R\} \simeq R$. Also f is onto, since any element (α, β) of R^2 is f (0, α, β). Thus, by Theorem 13, $R^3/R \simeq R^2$.

Note : In general, for any $n \geq m$, $R^n/R^m \simeq R^{n-m}$. Similarly, $C^{n-m} \simeq C^n/C^m$ for $n \geq m$.

The next result is a corollary to the Fundamental Theorem of Homomorphism.

But, before studying it, read Unit 3 for the definition of the sum of spaces.

Corollary 1: Let A and B be subspaces of a vector space V. Then A + B/B $\simeq$ A/A $\cap$ B.

Proof: We define a linear function $T : A \to \frac{A+B}{B}$ by T (a) = a + B.

T is well defined because a + B is an element of $\frac{A+B}{B}$ (since a = a + 0 $\in$ A + B).

T is a linear transformation because, for α_1, α_2 in F and a_1, a_2 in A, we have

$$T(\alpha_1 a_1 + \alpha_2 a_2) = \alpha_1 a_1 + \alpha_2 a_2 + B = \alpha_1 (a_1 + B) + \alpha_2 (a_2 + B)$$
$$= \alpha_1 T(a_1) + \alpha_2 T(a_2).$$

Now we will show that T is surjective. Any element of $\frac{A+B}{B}$ is of the form a + b + B, where a $\in$ A and b $\in$ B.

Now a + b + B = a + B + b + B = a + B + B, since b $\in$ B,

= a + B, since B is the zero element of $\frac{A+B}{B}$

= T (a), proving that T is surjective.

$\therefore R(T) = \frac{A+B}{B}$

We will now prove that Ker T = A $\cap$ B.

If $a \in$ Ker T, then $a \in A$ and T (a) = 0. This means that a + B = B, the zero element of $\frac{A+B}{B}$
Hence, $a \in B$ (by Unit 3, E23). Therefore, $a \in A \cap B$. Thus, Ker T $\subseteq A \cap B$. On the other hand, $a \in A \cap B \Rightarrow a \in A$ and $a \in B \Rightarrow a \in A$ and $a + B = B \Rightarrow a \in A$ and
T (a) = T (0) = 0

$\Rightarrow a \in$ Ker T.

This proves that $A \cap B$ = Ker T.

Now using Theorem 13, we get

A/Ker T $\simeq$ R (T).

That is, $A/(A \cap B) \simeq (A + B)/B$

E 26) Using the corollary above, show that $A \oplus B/B \simeq A$ ($\oplus$ denotes the direct sum defined in Sec. 3.6).

There is yet another interesting corollary to the Fundamental Theorem of Homomorphism.

Corollary 2: Let W be a subspace of a vector space V. Then, for any subspace U of V containing W,

$$\frac{V/W}{U/W} \simeq V/U$$

Proof: This time we shall prove the theorem with you. To start with let us define a function T : V/W → V/U : T (v + W) = v + U. Now try E 27.

E 27) a) Check that T is well defined.

b) Prove that T is a linear transformation.

c) What are the spaces Ker T and R (T)?

So, is the theorem proved? Yes; apply Theorem 13 to T.

We end the unit by summarising what we have done in it.

5.6 SUMMARY

In this unit we have covered the following points.

1) A linear transformation from a vector space U over F to a vector space V over F is a function T: U → V such that,

LT1) $T(u_1 + u_2) = T(u_1) + T(u_2) \; \forall u_1, u_2 \in U$, and

LT2) $T(\alpha u) = \alpha T(u)$, for $\alpha \in F$ and $u \in U$.

These conditions are equivalent to the single condition

LT3) $T(\alpha u_1 + \beta u_2) = \alpha T(u_1) + \beta T(u_2)$ for $\alpha, \beta \in F$ and $u_1, u_2 \in U$.

2) Given a linear transformation $T : U \to V$,

   i) the kernel of T is the vector space $\{u \in U \mid T(u) = 0\}$, denoted by Ker T.

   ii) the range of T is the vector space $\{T(u) \mid u \in U\}$, denoted by R(T).

   iii) The rank of T = $\dim_F R(T)$.

   iv) The nullity of T = $\dim_F \text{Ker}\, T$.

3) Let U and V be finite-dimensional vector spaces over F and $T: U \to V$ be a linear transformation. Then rank (T) + nullity (T) = dim U.

4) Let $T: U \to V$ be a linear transformation. Then

   i) T is one-one if $T(u_1) = T(u_2) \Rightarrow u_1 = u_2 \; \forall u_1, u_2 \in U$.

   ii) T is onto if, for any $v \in V \; \exists \; u \in U$ such that $T(u) = v$.

   iii) T is an isomorphism (or, is invertible) if it is one-one and onto, and then U and V are called isomorphic spaces. This is denoted by $U \simeq V$.

5) $T : U \to V$ is

   i) one-one if and only if Ker T = {0}

   ii) onto if and only if R (T) = V.

6) Let U and V be finite-dimensional vector spaces with the same dimension. Then $T: U \to V$ is 1 – 1 iff T is onto iff T is an isomorphism.

7) Two finite-dimensional vector spaces U and V are isomorphic if and only if dim U = dim V

8) Let V and W be vector spaces over a field F, and $T : V \to W$ be a linear transformation. Then $V/\text{Ker}\, T \simeq R(T)$.

5.7 SOLUTIONS/ANSWERS

E1) For any $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$, we know that $\alpha_1 u_1 \in U$ and $\alpha_2 u_2 \in U$. Therefore, by LT1,

$$T(\alpha_1 u_1 + \alpha_2 u_2) = T(\alpha_1 u_1) + T(\alpha_2 u_2)$$
$$= \alpha_1 T(u_1) + \alpha_2 T(u_2), \text{ by LT2.}$$

Thus, LT3 is true.

E 2) By LT2, $T(0.u) = 0.T(u)$ for any $u \in U$. Thus, $T(0) = 0$. Similarly, for any $u \in U$, $T(-u) = T((-1)u) = (-1)T(u) = -T(u)$.

E 3) $T(x, y) = (-x, y) \; \forall (x, y) \in R^2$. (See the geometric view in Fig. 4.) T is a linear operator. This can be proved the same way as we did in Example 4.

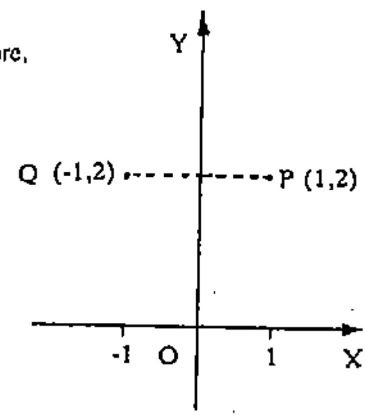


Fig. 4: Q is the reflection of P in the y-axis.

E 4) $T((x_1, x_2, x_3) + (y_1, y_2, y_3)) = T(x_1 + y_1, x_2 + y_2, x_3 + y_3)$

$= a_1(x_1 + y_1) + a_2(x_2 + y_2) + a_3(x_3 + y_3)$

$= (a_1 x_1 + a_2 x_2 + a_3 x_3) + (a_1 y_1 + a_2 y_2 + a_3 y_3)$

$= T(x_1, x_2, x_3) + T(y_1, y_2, y_3)$

Also, for any $\alpha \in R$,

$T(\alpha(x_1, x_2, x_3)) = a_1 \alpha x_1 + a_2 \alpha x_2 + a_3 \alpha x_3$

$= \alpha(a_1 x_1 + a_2 x_2 + a_3 x_3) = \alpha T(x_1, x_2, x_3)$.

Thus, LT1 and LT2 hold for T.

E 5) We will check if LT1 and LT2 hold. Firstly,

$T((x_1, x_2, x_3) + (y_1, y_2, y_3)) = T(x_1 + y_1, x_2 + y_2, x_3 + y_3)$

$= (x_1 + y_1 + x_2 + y_2 - x_3 - y_3, 2x_1 + 2y_1 - x_2 - y_2, x_2 + y_2 + 2x_3 + 2y_3)$

$= (x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3) + (y_1 + y_2 - y_3, 2y_1 - y_2, y_2 + 2y_3)$

$= T(x_1, x_2, x_3) + T(y_1, y_2, y_3)$, showing that LT1 holds.

Also, for any $\alpha \in \mathbf{R}$,

$T(\alpha(x_1, x_2, x_3)) = T(\alpha x_1, \alpha x_2, \alpha x_3)$

$= (\alpha x_1 + \alpha x_2 - \alpha x_3, 2\alpha x_1 - \alpha x_2, \alpha x_2 + 2\alpha x_3)$

$= \alpha(x_1 + x_2 - x_3, 2x_1 - x_2, x_2 + 2x_3) = \alpha T(x_1, x_2, x_3)$, showing that LT2 holds.

E 6) We want to show that $D(\alpha f + \beta g) = \alpha D(f) + \beta D(g)$, for any $\alpha, \beta \in \mathbf{R}$ and $f, g \in \mathbf{P}_n$.
Now, let $f(x) = a_0 + a_1 x + a_2 x^2 + .. + a_n x^n$ and $g(x) = b_0 + b_1 x + ... + b_n x^n$.

Then $(\alpha f + \beta g)(x) = (\alpha a_0 + \beta b_0) + (\alpha a_1 + \beta b_1) x + ... + (\alpha a_n + \beta b_n) x^n$.

$\therefore [D(\alpha f + \beta g)](x) = (\alpha a_1 + \beta b_1) + 2(\alpha a_2 + \beta b_2) x + ... + n(\alpha a_n + \beta b_n) x^{n-1}$

$= \alpha(a_1 + 2a_2 x + ... + na_n x^{n-1}) + \beta(b_1 + 2b_2 x + ... + nb_n x^{n-1})$

$= \alpha(Df)(x) + \beta(Dg)(x) = (\alpha Df + \beta Dg)(x)$

Thus, $D(\alpha f + \beta g) = \alpha Df + \beta Dg$, showing that D is a linear map.

E 7) No. Because, if T exists, then

$T(2u_1 + u_2) = 2T(u_1) + T(u_2)$.

But $2u_1 + u_2 = u_3$. $\therefore T(2u_1 + u_2) = T(u_3) = v_3 = (1, 1)$.

On the other hand, $2T(u_1) + T(u_2) = 2v_1 + v_2 = (2,0) + (0,1)$

$= (2,1) \neq v_3$.

Therefore, LT3 is violated. Therefore, no such T exists.

E 8) Note that $\{(1,0), (0,5)\}$ is a basis for $\mathbf{R}^2$.

Now $(3,5) = 3(1,0) + (0,5)$.

Therefore, $T(3,5) = 3T(1,0) + T(0,5) = 3(0,1) + (1,0) = (1,3)$.

Similarly, $(5,3) = 5(1,0) + 3/5(0,5)$.

Therefore, $T(5,3) = 5(0,1) + 3/5(1,0) = (3/5, 5)$.

Note that $T(5,3) \neq T(3,5)$

E 9) a) $\dim_{\mathbf{R}} \mathbf{C} = 2$, a basis being $\{1, i\}$, $i = \sqrt{-1}$.

b) Let $T: \mathbf{C} \to \mathbf{R}$ be such that $T(1) = \alpha$, $T(i) = \beta$.

Then, for any element $x + iy \in \mathbf{C}$ $(x, y \in \mathbf{R})$, we have $T(x + iy) = xT(1) + yT(i) = x\alpha + y\beta$. Thus, T is defined by $T(x + iy) = x\alpha + y\beta \; \forall x + iy \in \mathbf{C}$.

E 10) $T: U \to V: T(u) = 0 \; \forall u \in U$.

$\therefore$, $\text{Ker } T = \{u \in U \mid T(u) = 0\} = U$

$R(T) = \{T(u) \mid u \in U\} = \{0\}$. $\therefore 1 \notin R(T)$.

E 11) a) $R(T) = \{T(x, y, z) \mid (x, y, z) \in \mathbf{R}^3\} = \{(x, y) \mid (x,y,z) \in \mathbf{R}^3\} = \mathbf{R}^2$.

$\text{Ker } T = \{(x, y, z) \mid T(x,y,z) = 0\} = \{(x, y, z) \mid (x, y) = (0, 0)\}$

$= \{(0,0,z) \mid z \in \mathbf{R}\}$

$\therefore$, Ker T is the z-axis.

b) $R(T) = \{z \mid (x, y, z) \in \mathbf{R}^3\} = \mathbf{R}$.

$\text{Ker } T = \{(x,y,0) \mid x, y \in \mathbf{R}\} = xy$ – plane in $\mathbf{R}^3$.

c) $R(T) = \{(x, y, z) \in \mathbf{R}^3 \mid \exists x_1, x_2, x_3 \in \mathbf{R}$ such that $x = x_1 + x_2 + x_3 = y = z\}$

$= \{(x,x,x) \in \mathbf{R}^3 \mid x = x_1 + x_2 + x_3$ for some $x_1, x_2, x_3 \in \mathbf{R}\}$

$= \{(x,x,x) \in \mathbf{R}^3 \mid x \in \mathbf{R}\}$

because, for any $x \in \mathbf{R}$, $(x, x, x) = T(x, 0, 0)$

$\therefore$, R(T) is generated by $\{(1,1,1)\}$.

$\text{Ker } T = \{(x_1, x_2, x_3) \mid x_1 + x_2 + x_3 = 0\}$, which is the plane $x_1 + x_2 + x_3 = 0$, in $\mathbf{R}^3$.

E 12) Any element of R(T) is of the form T(u), $u \in U$. Since $\{e_1, ..., e_n\}$ generates U, $\exists$ scalars $\alpha_1, ..., \alpha_n$ such that $u = \alpha_1 e_1 + ... + \alpha_n e_n$.

Then $T(u) = \alpha_1 T(e_1) + ... + \alpha_n T(e_n)$, that is, T(u) is in the linear span of $\{T(e_1), ..., T(e_n)\}$.

$\therefore \{T(e_1), ..., T(e_n)\}$ generates R(T).

E13) $T: V \to V : T(v) = v$. Since $R(T) = V$ and $\text{Ker } T = \{0\}$, we see that rank $(T) = \dim V$, nullity $(T) = 0$.

E14) $R(D) = \{a_1 + 2a_2x + \ldots + na_nx^{n-1} \mid a_1, \ldots, a_n \in \mathbf{R}\}$

Thus, $R(D) \subseteq P_{n-1}$. But any element $b_0 + b_1x + \ldots + b_{n-1}x^{n-1}$, in

P_{n-1} is $D\left(b_0x + \frac{b_1}{2}x^2 + \cdots + \frac{b_{n-1}}{n}x^n\right) \in R(D)$.

Therefore, $R(D) = P_{n-1}$.

$\therefore$, a basis for $R(D)$ is $\{1, x, \ldots, x^{n-1}\}$, and rank $(D) = n$.

$\text{Ker } D = \{a_0 + a_1x + \ldots + a_nx^n \mid a_1 + 2a_2x + \ldots + na_nx^{n-1} = 0, a_i \in \mathbf{R} \forall i\}$

$= \{a_0 + a_1x + \ldots + a_nx^n \mid a_1 = 0, a_2 = 0, \ldots, a_n = 0, a_i \in \mathbf{R} \forall i\}$

$= \{a_0 \mid a_0 \in \mathbf{R}\} = \mathbf{R}$.

$\therefore$, a basis for Ker D is $\{1\}$.

$\Rightarrow$ nullity $(D) = 1$.

E 15) a) We have shown that $R(T) = \mathbf{R}^2$. $\therefore$ rank $(T) = 2$.

Therefore, nullity $(T) = \dim \mathbf{R}^3 - 2 = 1$.

b) rank $(T) = 1$, nullity $(T) = 2$.

c) $R(T)$ is generated by $\{(1, 1, 1)\}$. $\therefore$ rank $(T) = 1$.

$\therefore$ nullity $(T) = 2$.

E 16) Now rank (T) + nullity $(T) = \dim U = 1$.

Also rank $(T) \geq 0$, nullity $(T) \geq 0$.

$\therefore$, the only values rank (T) can take are 0 and 1. If rank $(T) = 0$, then $\dim R(T) = 0$. Thus, $R(T) = \{0\}$, that is, $R(T)$ is a point.

If rank $(T) = 1$, then $\dim R(T) = 1$. That is, $R(T)$ is a vector space over $\mathbf{R}$ generated by a single element, v, say. Then $R(T)$ is the line $\mathbf{R}_v = \{\alpha v \mid \alpha \in \mathbf{R}\}$.

E 17) By Theorem 5, nullity $(ST) = \dim V -$ rank (ST). By (a) of Theorem 6, we know that $-$ rank $(ST) \geq -$ rank (S) and $-$ rank $(ST) \geq -$ rank (T).

$\therefore$, nullity $(ST) \geq \dim V -$ rank (S) and nullity $(ST) \geq \dim V -$ rank (T).

Thus, nullity $(ST) \geq$ nullity (S) and nullity $(ST) \geq$ nullity (T). That is, nullity $(ST) \geq \max$ {nullity (S), nullity (T)}.

E18) Since $1 \neq 2$, but $0(1) = 0(2) = 0$, we find that 0 is not $1 - 1$.

E19) Firstly note that T is a linear transformation. Secondly, T is $1 - 1$ because $T(x, y, z) = T(x', y', z') \Rightarrow (x, y, z) = (x', y', z')$

Thirdly, T is onto because any $(x,y,z) \in \mathbf{R}^3$ can be written as $T(x - y, y, z)$

$\therefore$, T is an isomorphism. $\therefore$ $T^{-1}: \mathbf{R}^3 \to \mathbf{R}^3$ exists and is defined by $T^{-1}(x, y, z) = (x - y, y, z)$.

E20) T is not an isomorphism because T is not $1 - 1$, since $(1, -1, 1) \in \text{Ker } T$.

E21) The linear operator in E11) (a) suffices.

E22) a) Let $\alpha_1, \ldots, \alpha_k \in F$ such that $\alpha_1 T(u_1) + \ldots + \alpha_k T(u_k) = 0$.

$\Rightarrow T(\alpha_1u_1 + \ldots + \alpha_ku_k) = 0 = T(0)$

$\Rightarrow \alpha_1u_1 + \ldots + \alpha_ku_k = 0$, since T is $1 - 1$.

$\Rightarrow \alpha_1 = 0, \ldots, \alpha_k = 0$, since $\{u_1, \ldots, u_k\}$ is linearly independent

$\therefore \{T(u_1), \ldots, T(u_k)\}$ is linearly independent.

b) No. For example, the zero operator maps every linearly independent set to $\{0\}$, which is not linearly independent.

c) Let $T: U \to V$ be a linear operator, and $\{u_1, \ldots, u_n\}$ be a linearly dependent set of vectors in U. We have to show that $\{T(u_1), \ldots, T(u_n)\}$ is linearly dependent. Since $\{u_1, \ldots, u_n\}$ is linearly dependent, $\exists$ scalars $a_1, \ldots, a_n$, not all zero, such that

$a_1u_1 + \ldots + a_nu_n = 0$.

Then $a_1 T(u_1) + \ldots + a_n T(u_n) = T(0) = 0$, so that $\{T(u_1), \ldots, T(u_n)\}$ is linearly dependent.

E23) T is a linear transformation. Now, if $(x, y, z) \in \text{Ker } T$, then $T(x, y, z) = (0, 0, 0)$.

$\therefore x + y = 0 = y + z = x + z \Rightarrow x = 0 = y = z$

$\Rightarrow \text{Ker } T = \{(0, 0, 0)\}$

$\Rightarrow$ T is 1 – 1.

$\therefore$, by Theorem 10, T is invertible.

To define $T^{-1}: R^3 \to R^3$ suppose $T^{-1}(x, y, z) = (a, b, c)$.

Then $T(a, b, c) = (x, y, z)$

$\Rightarrow (a + b, b + c, a + c) = (x, y, z)$

$\Rightarrow a + b = x, b + c = y, a + c = z$

$$\Rightarrow a = \frac{x + z - y}{2},\ b = \frac{x + y - z}{2},\ c = \frac{y + z - x}{2}$$

$$\therefore T^{-1}(x, y, z) = \left(\frac{x + z - y}{2}, \frac{x + y - z}{2}, \frac{y + z - x}{2}\right) \text{ for any } (x, y, z) \in \mathbf{R}^3.$$

E24) $T : U \to V$ is 1 – 1. **Suppose T is onto. Then T is an isomorphism and dim U = dim V,** by Theorem 12. Conversely, suppose dim U = dim V. Then T is onto by Theorem 10.

E25) The Rank Nullity Theorem and Example 14 give

dim V – nullity (ST) = dim V – nullity (T) – dim (R (T) $\cap$ Ker S)

$\Rightarrow$ nullity (ST) = nullity (T) + dim (R (T) $\cap$ Ker S)

E26) In the case of the direct sum $A \oplus B$, we have $A \cap B = \{0\}$.

$$\therefore \frac{A \oplus B}{B} = A$$

E27) a) $v + W = v' + W \Rightarrow v - v' \in W \subseteq U \Rightarrow v - v' \in U \Rightarrow v + U = v' + U$

$\Rightarrow T(v + W) = T(v' + W)$

$\therefore$ T is well defined.

b) For any $v + W$, $v' + W$ in V/W and scalars a, b, we have

$T(a(v + W) + b(v' + W)) = T(av + bv' + W) = av + bv' + U$

$= a(v + U) + b(v' + U) = aT(v + W) + bT(v' + W)$.

$\therefore$, T is a linear operator.

c) Ker T = $\{v + W \mid v + U = U\}$, since U is the "zero" for V/U.

$= \{v + W \mid v \in U\} = U/W$.

$R(T) = \{v + U \mid v \in V\} = V/U$.

UNIT 6 LINEAR TRANSFORMATIONS-II

Structure

6.1 INTRODUCTION

In the last unit we introduced you to linear transformations and their properties. We will now show that the set of all linear transformations from a vector space U to a vector space V forms a vector space itself, and its dimension is (dim U) (dim V). In particular, we define and discuss the dual space of a vector space.

In Unit 1 we defined the composition of two functions. Over here we will discuss the composition of two linear transformations and show that it is again a linear operator. Note that we use the terms 'linear transformation' and 'linear operator' interchangeably.

Finally, we study polynomials with coefficients from a field F, in a linear operator T : V → V. You will see that every such T satisfies a polynomial equation g(x) = 0. That is, f we substitute T for x in g(x) we get the zero transformation. We will, then, define the ninimal polynomial of an operator and discuss some of its properties. These ideas will crop ıp again in Unit 11.

ou must revise Units 1 and 5 before going further.

)bjectives

\fter reading this unit, you should be able to

- prove and use the fact that L (U, V) is a vector space of dimension (dim U) (dim V);
- use dual bases, whenever convenient;
- obtain the composition of two linear operators, whenever possible;
- obtain the minimal polynomial of a linear transformation T : V → V in some simple cases;
- obtain the inverse of an isomorphism T: V → V if its minimal polynomial is known.

.2 THE VECTOR SPACE L (U, V)

y now you must be quite familiar with linear operators, as well as vector spaces. In this ction we consider the set of all linear operators from one vector space to another, and ıow that it forms a vector space.

et U, V be vector spaces over a field F. Consider the set of all linear transformations from to V. We denote this set by L (U, V).

'e will now define addition and scalar multiplication in L (U, V) so that L (U, V) becomes vector space.

ıppose S, T ∈ L (U, V) (that is, S and T are linear operators from U to V). We define + T) : U → V by

+ T) (u) = S (u) + T (u) ∀u ∈ U.

ɔw, for α_1, α_2 ∈ F and u_1, u_2 ∈ U, we have

$(S + T)(a_1 u_1 + a_2 u_2)$

$= S(a_1 u_1 + a_2 u_2) + T(a_1 u_1 + a_2 u_2)$

$= a_1 S(u_1) + a_2 S(u_2) + a_1 T(u_1) + a_2 T(u_2)$

$= a_1 (S(u_1) + T(u_1)) + a_2 (S(u_2) + T(u_2))$

$= a_1 (S + T)(u_1) + a_2 (S + T)(u_2)$

Hence, $S + T \in L(U, V)$.

Next, suppose $S \in L(U, V)$ and $\alpha \in F$. We define $\alpha S: U \to V$ as follows:

$(\alpha S)(u) = \alpha S(u) \;\forall u \in U.$

Is αS a linear operator? To answer this take $\beta_1, \beta_2 \in F$ and $u_1, u_2 \in U$. Then,

$$(\alpha S)(\beta_1 u_1 + \beta_2 u_2) = \alpha S(\beta_1 u_1 + \beta_2 u_2) = \alpha[\beta_1 S(u_1) + \beta_2 S(u_2)]$$
$$= \beta_1 (\alpha S)(u_1) + \beta_2 (\alpha S)(u_2)$$

Hence, $\alpha S \in L(U, V)$.

So we have successfully defined addition and scalar multiplication on L (U,V).

E E1) Show that the set L (U,V) is a vector space over F with respect to the operations of addition and multiplication by scalars defined above. (Hint: The zero vector in this space is the zero transformation. All the conditions VS1 – VS10 (of Unit 3) have to be verified.)

Notation: For any vector space V we denote **L(V,V) by A(V)**.

Let U and V be vector spaces over F of dimensions m and n, respectively. We have already observed that L(U,V) is a vector space over F. Therefore, it must have a dimension. We now show that the dimension of L(U,V) is mn.

Theorem 1: Let U,V be vector spaces over a field F of dimensions m and n, respectively. Then L(U,V) is a vector space of dimension mn.

Proof: Let $\{e_1, \ldots, e_m\}$ be a basis of U and $\{f_1, \ldots, f_n\}$ be a basis of V. By Theorem 3 of Unit 5, there exists a unique linear transformation $E_{11} \in L(U, V)$, such that

$E_{11}(e_1) = f_1, E_{11}(e_2) = 0, \ldots, E_{11}(e_m) = 0.$

Similarly, $E_{12} \in L(U, V)$ such that

$E_{12}(e_1) = 0, E_{12}(e_2) = f_1, E_{12}(e_3) = 0, \ldots, E_{12}(e_m) = 0.$

In general, there exist $E_{ij} \in L(U, V)$ for $i = 1, \ldots, n$, $j = 1, \ldots, m$, such that $E_{ij}(e_j) = f_i$ and $E_{ij}(e_k) = 0$ for $j \neq k$.

To get used to these E_{ij} try the following exercise before continuing the proof.

E2) Clearly define E_{2m}, E_{32} and E_{mn}.

Now, let us go on with the proof of Theorem 1.

If $u = c_1 e_1 + \ldots + c_m e_m$, where $c_i \in F \ \forall i$, then $E_{ij}(u) = c_j f_i$.

We complete the proof by showing that $\{E_{ij} \mid i = 1, \ldots, n, j = 1, \ldots, m\}$ is a basis of L(U, V).

Let us first show that this set is linearly independent over F. For this, suppose

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij} = 0 \qquad \ldots\ldots (1)$$

where $c_{ij} \in F$. We must show that $c_{ij} = 0$ for all i, j.

(1) implies that

$$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij}(e_k) = 0 \ \forall \ k = 1, \ldots, m.$$

Thus, by definition of the E_{ij}'s, we get

$$\sum_{i=1}^{n} c_{ik} f_i = 0,$$

But $\{f_1, \ldots, f_n\}$ is a basis for V. Thus, $c_{ik} = 0$, for all $i = 1, \ldots, n$.

But this is true for all $k = 1, \ldots, m$.

Hence, we conclude that $c_{ij} = 0 \ \forall i, j$. Therefore, the set of E_{ij}'s is linearly independent.

Next, we show that the set $\{E_{ij} \mid i = 1, \ldots, n, j = 1, \ldots, m\}$ spans L (U, V). Suppose $T \in L(U, V)$.

Now, for each j such that $1 \leq j \leq m$, $T(e_j) \in V$. Since $\{f_1, \ldots, f_n\}$ is a basis of V, there exist scalars $c_{1j}, \ldots, c_{nj}$ such that

$$T(e_j) = \sum_{i=1}^{n} c_{ij} f_i \qquad \ldots\ldots (2)$$

We shall prove that

$$T = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij} \qquad \ldots\ldots (3)$$

By Theorem 1 of Unit 5 it is enough to show that, for each k with $1 \leq k \leq m$,

$$T(e_k) = \sum_i \sum_j c_{ij} E_{ij}(e_k)$$

Now,

$\sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} E_{ij}(e_k) = \sum_{i=1}^{n} c_{ik} f_i = T(e_k)$, by (2). This implies (3).

Thus, we have proved that the set of mn elements $\{E_{ij} \mid i = 1, \ldots, n, j = 1, \ldots, m\}$ is a basis for L(U, V).

Let us see some ways of using this theorem.

Example 1: Show that $L(\mathbf{R}^2, \mathbf{R})$ is a plane.

Solution: $L(\mathbf{R}^2, \mathbf{R})$ is a real vector space of dimension $2 \times 1 = 2$.

Thus, by Theorem 12 of Unit 5 $L(\mathbf{R}^2, \mathbf{R}) \simeq \mathbf{R}^2$, the real plane.

Example 2: Let U, V be vector spaces of dimensions m and n, respectively. Suppose W is a subspace of V of dimension p ($\leq$ n). Let

$X = \{T \in L(U, V): T(u) \in W \text{ for all } u \in U\}$

Is X a subspace of L (U, V)? If yes, find its dimension.

Solution: $X = \{T \in L(U, V) \mid T(U) \subseteq W\} = L(U, W)$. Thus, X is also a vector space. Since it is a subset of L(U, V), it is a subspace of L(U, V). By Theorem 1, dim X = mp.

E E3) What can be a basis for $L(\mathbf{R}^2, \mathbf{R})$, and for $L(\mathbf{R}, \mathbf{R}^2)$? Notice that both these spaces have the same dimension over R.

After having looked at L(U, V), we now discuss this vector space for the particular case when V = F.

6.3 THE DUAL SPACE

The vector space L(U, V), discussed in Sec. 6.2, has a particular name when V = F.

Recall that F is also a vector space over F.

Definition: Let U be a vector space over F. Then the space L(U, F) is called the **dual space** of U, and is denoted by U*.

In this section we shall study some basic properties of U*.

The elements of U* have a specific name, which we now give.

Definition: A linear transformation $T: U \to F$ is called a **linear functional.**

Thus, a linear functional on U is a function $T: U \to F$ such that $T(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 T(u_1) + \alpha_2 T(u_2)$, for $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$.

For example, the map $f: \mathbf{R}^3 \to \mathbf{R}: f(x_1, x_2, x_3) = a_1 x_1 + a_2 x_2 + a_3 x_3$, where $a_1, a_2, a_3 \in \mathbf{R}$ are fixed, is a linear functional on $\mathbf{R}^3$. You have already seen this in Unit 5 (E4).

E E4) Prove that any linear functional on $\mathbf{R}^3$ is of the form given in the example above.

We now come to a very important aspect of the dual space.

We know that the space V*, of linear functionals on V, is a vector space. Also, if dim V = m, then dim V* = m, by Theorem 1. (Remember, dim F = 1.)

Hence, we see that **dim V = dim V***. From Theorem 12 of Unit 5, it follows that the vector spaces V and V* are isomorphic.

We now construct a special basis for V*. Let $\{e_1, ..., e_m\}$ be a basis of V. By Theorem 3 of Unit 5, for each i = 1, ... m, there exists a unique linear functional f_i on V such that

$$f_i(e_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$
$$= \delta_{ij}$$

The Kronecker delta function is
$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

We will prove that the linear functionals $f_1,...,f_m$, **constructed above, form a basis of V*.**
Since dim V = dim V* = m, it is enough to show that the set $\{f_1, ..., f_m\}$ is linearly independent. For this we suppose $c_1,...,c_m \in$ **F such that** $c_1 f_1 + ... + c_m f_m = 0$.

We must show that $c_i = 0$, for all i.

Now $\sum_{j=1}^{n} c_j f_j = 0$

$$\Rightarrow \left(\sum_{j=1}^{n} c_j f_j\right)(e_i) = 0, \quad \text{for each } i.$$

$$\Rightarrow \sum_{j=1}^{n} c_j (f_j(e_i)) = 0 \ \forall \ i$$

$$\Rightarrow \sum_{j=1}^{n} c_j \delta_{ji} = 0 \ \forall \ i \Rightarrow c_i = 0 \ \forall \ i.$$

Thus, the set $\{f_1, ..., f_m\}$ is a set of m linearly independent elements of a vector space V* of dimension m. Thus, from Unit 4 (Theorem 5, Cor. 1), it forms a basis of V*.

Definition: The basis $\{f_1,...., f_m\}$ of V* is called the **dual basis** of the basis $\{e_1,...., e_m\}$ of V.

We now come to the result that shows the convenience of using a dual basis.

Theorem 2: Let V be a vector space over F of dimension n, $\{e_1,...,e_n\}$ be a basis of V and $\{f_1,...,f_n\}$ be the dual basis of $\{e_1,....,e_n\}$. Then, for each $f \in V^*$,

$$f = \sum_{i=1}^{n} f(e_i) f_i$$

and, for each $v \in V$,

$$v = \sum_{i=1}^{n} f_i(v) e_i.$$

Proof: Since $\{f_1, ..., f_n\}$ **is a basis of** V*, for $f \in V^*$ there exist scalars $c_1,....,c_n$ such that

$$f = \sum_{i=1}^{n} c_i f_i.$$

Therefore,

$$f(e_j) = \sum_{i=1}^{n} c_i f_i(e_j)$$
$$= \sum_{i=1}^{n} c_i \delta_{ij}, \text{ by definition of dual basis.}$$
$$= c_j.$$

This implies that $c_i = f(e_i) \forall i = 1, ..., n$. Therefore, $f = \sum_{i=1}^{n} f(e_i) f_i$. Similarly, for $v \in V$, there exist scalars $a_1,...,a_n$ such that

$$v = \sum_{i=1}^{n} a_i e_i$$

Hence,
$$f_j(v) = \sum_{i=1}^{n} a_i f_j(e_i)$$
$$= \sum_{i=1}^{n} a_i \delta_{ji}$$
$$= a_j,$$

and we obtain

$$v = \sum_{i=1}^{n} f_i(v)e_i.$$

Let us see an example of how this theorem works.

Example 3: Consider the basis $e_1 = (1, 0, -1)$, $e_2 = (1, 1, 1)$, $e_3 = (1, 1, 0)$ of C^3 over C. Find the dual basis of $\{e_1, e_2, e_3\}$.

Solution: Any element of C^3 is $v=(z_1, z_2, z_3)$, $z_i \in C$. Since $\{e_1, e_2, e_3\}$ is a basis, we have $\alpha_1, \alpha_2, \alpha_3 \in C$. Since that

$$v = (z_1, z_2, z_3) = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$$

$$= (\alpha_1 + \alpha_2 + \alpha_3, \alpha_2 + \alpha_3, -\alpha_1 + \alpha_2)$$

Thus, $\alpha_1 + \alpha_2 + \alpha_3 = z_1$

$$\alpha_2 + \alpha_3 = z_2$$

$$-\alpha_1 + \alpha_2 = z_3$$

These equations can be solved to get

$$\alpha_1 = z_1 - z_2,\ \alpha_2 = z_1 - z_2 + z_3,\ \alpha_3 = 2z_2 - z_1 - z_3$$

Now, by Theorem 2,

$v = f_1(v)e_1 + f_2(v)e_2 + f_3(v)e_3$, where $\{f_1, f_2, f_3\}$ is the dual basis. Also $v = \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3$.

Hence, $f_1(v) = \alpha_1$, $f_2(v) = \alpha_2$, $f_3(v) = \alpha_3$ $\forall v \in C^3$.

Thus, the dual basis of $\{e_1, e_2, e_3\}$ is $\{f_1, f_2, f_3\}$, where f_1, f_2, f_3 will be defined as follows:

$$f_1(z_1, z_2, z_3) = \alpha_1 = z_1 - z_2$$

$$f_2(z_1, z_2, z_3) = \alpha_2 = z_1 - z_2 + z_3$$

$$f_3(z_1, z_2, z_3) = \alpha_3 = 2z_2 - z_1 - z_3$$

E E5) What is the dual basis for the basis $\{1, x, x^2\}$ of the space

$P_2 = \{a_0 + a_1 x + a_2 x^2 \mid a_i \in R\}$?

Now let us look at the dual of the dual space. If you like, you may skip this portion and go straight to Sec. 6.4.

Let V be an n-dimensional vector space. We have already seen that V and V* are isomorphic because dim V = dim V*. The dual of V* is called the **second dual of V** and is denoted by V**. We will show that $V \simeq V^{**}$

Now any element of V** is a linear transformation from V* to F. Also, for any $v \in V$ and $f \in V^*$, $f(v) \in F$. So we define a mapping $\phi : V \to V^{**} : v \to \phi v$, where $(\phi v)(f) = f(v)$ for all $f \in V^*$ and $v \in V$. (Over here we will use $\phi(v)$ and ϕv interchangeably.)

Note that, for any $v \in V$, ϕv is a well defined mapping from $V^* \to F$. We have to check that it is a linear mapping.

Now, for $c_1, c_2 \in F$ and $f_1, f_2 \in V^*$,

$$(\phi v)(c_1 f_1 + c_2 f_2) = (c_1 f_1 + c_2 f_2)(v)$$

$$= c_1 f_1(v) + c_2 f_2(v)$$

$$= c_1 (\phi v)(f_1) + c_2 (\phi v)(f_2)$$

$\therefore \phi v \in L(V^*, F) = V^{**}$ $\forall v$.

Furthermore, the map $\phi : V \to V^{**}$ is linear. This can be seen as follows: for $c_1, c_2 \in F$ and $v_1, v_2 \in V$.

$$\begin{aligned}\phi (c_1 v_1 + c_2 v_2) (f) &= f(c_1 v_1 + c_2 v_2)\\ &= c_1 f(v_1) + c_2 f(v_2)\\ &= c_1 (\phi v_1)(f) + c_2 (\phi v_2)(f)\\ &= (c_1 \phi v_1 + c_2 \phi v_2)(f).\end{aligned}$$

This is true $\forall f \in V^*$. Thus, $\phi (c_1 v_1 + c_2 v_2) = c_1 \phi (v_1) + c_2 \phi (v_2)$.

Now that we have shown that ϕ is linear, we want to show that it is actually an isomorphism. We will show that ϕ is $1-1$. For this, by Theorem 7 of Unit 5, it suffices to show that $\phi (v) = 0$ implies $v = 0$. Let $\{f_1, \ldots, f_n\}$ be the dual basis of a basis $\{e_1, \ldots, e_n\}$ of V.

By Theorem 2, we have $v = \sum_{i=1}^{n} f_i(v) e_i$.

$$\begin{aligned}\text{Now } \phi (v) = 0 &\Rightarrow (\phi v)(f_i) = 0 \ \forall i = 1, \ldots, n\\ &\Rightarrow f_i(v) = 0 \ \forall i = 1, \ldots, n\\ &\Rightarrow v = \Sigma f_i(v) e_i = 0\end{aligned}$$

Hence, it follows that ϕ is 1–1. Thus, ϕ is an isomorphism (Unit 5, Theorem 10).

What we have just proved is the following theorem.

Theorem 3: The map $\phi : V \to V^{**}$, defined by $(\phi v)(f) = f(v) \ \forall \ v \in V$ and $f \in V^*$, is an isomorphism.

We now give an important corollary to this theorem.

Corollary: Let ψ be a linear functional on V^* (i.e., $\psi \in V^{**}$).

Then there exists a unique $v \in V$ such that

$\psi(f) = f(v)$ for all $f \in V^*$.

'ψ' is the Greek letter 'psi'.

Proof: By Theorem 3, since ϕ is an isomorphism, it is onto and $1-1$. Thus, there exists a unique $v \in V$ such that $\phi (v) = \psi$. This, by definition, implies that

$\psi (f) = (\phi v)(f) = f(v)$ for all $f \in V^*$.

Using the second dual try to prove the following exercise.

E6) Show that each basis of V^* is the dual of some basis of V.

In the following section we look at the composition of linear operators, and the vector space A(V), where V is a vector space over F.

6.4 COMPOSITION OF LINEAR TRANSFORMATIONS

Do you remember the definition of the composition of functions, which you studied in Unit 1? Let us now consider the particular case of the composition of two linear transformations. Suppose $T: U \to V$ and $S: V \to W$ are two linear transformations. The composition of S and T is a function $S \circ T: U \to W$, defined by

$S \circ T(u) = S(T(u)) \ \forall\ u \in U$.

This is diagrammatically represented in Fig. 1.

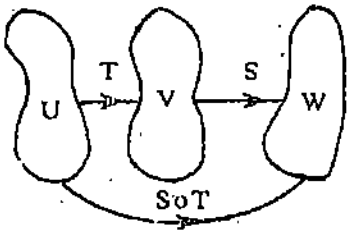


Fig. 1: SoT is the composition of S and T.

The first question which comes to our mind is whether SoT is linear. The affirmative answer is given by the following result.

Theorem 4: Let U, V, W be vector spaces over F. Suppose $S \in L(V, W)$ and $T \in L(U, V)$. Then $S \circ T \in L(U, W)$.

Proof: All we need to prove is the linearity of the map SoT. Let $\alpha_1, \alpha_2 \in F$ and $u_1, u_2 \in U$. Then

$S \circ T(\alpha_1 u_1 + \alpha_2 u_2) = S(T(\alpha_1 u_1 + \alpha_2 u_2))$

$= S(\alpha_1 T(u_1) + \alpha_2 T(u_2))$, since T is linear.

$= \alpha_1 S(T(u_1)) + \alpha_2 S(T(u_2))$, since S is linear

$= \alpha_1 S \circ T(u_1) + \alpha_2 S \circ T(u_2)$

This shows that $S \circ T \in L(U, W)$

Try the following exercises now.

E7) Let I be the identity operator on V. Show that $S \circ I = I \circ S = S$ for all $S \in A(V)$.

E8) Prove that $S \circ 0 = 0 \circ S = 0$ for all $S \in A(V)$, where 0 is the null operator.

We now make an observation.

Remark: Let $S: V \to V$ be an invertible linear transformation (ref. Sec. 5.4), that is, an isomorphism. Then, by Unit 5, Theorem 8, $S^{-1} \in L(V, V) = A(V)$.

Since $S^{-1} \circ S(v) = v$ and $S \circ S^{-1}(v) = v$ for all $v \in V$,

$S \circ S^{-1} = S^{-1} \circ S = I_V$, where I_V denotes the identity transformation on V.

This remark leads us to the following interesting result.

Theorem 5: Let V be a vector space over a field F. A linear transformation $S \in A(V)$ is an isomorphism if and only if $\exists\ T \in A(V)$ such that $S \circ T = I = T \circ S$.

Proof: Let us first assume that S is an isomorphism. Then, the remark above tells us that $\exists\ S^{-1} \in A(V)$ such that $S \circ S^{-1} = I = S^{-1} \circ S$. Thus, we have $T (= S^{-1})$ such that $S \circ T = T \circ S = I$.

Conversely, suppose T exists in A (V), such that $S \circ T = I = T \circ S$. We want to show that S is 1 – 1 and onto.

We first show that S is 1 – 1, that is, Ker S = {0}. Now, $x \in \text{Ker } S \Rightarrow S(x) = 0 \Rightarrow T \circ S(x) = T(0) = 0 \Rightarrow I(x) = 0 \Rightarrow x = 0$. Thus, Ker S = {0}.

Next, we show that S is onto, that is, for any $v \in V$, $\exists\ u \in V$ such that $S(u) = v$. Now, for any $v \in V$,

$v = I(v) = S \circ T(v) = S(T(v)) = S(u)$, where $u = T(v) \in V$. Thus, S is onto.

Hence, S is 1 – 1 and onto, that is, S is an isomorphism.

Use Theorem 5 to solve the following exercise.

E9) Let $S(x_1, x_2) = (x_2, -x_1)$ and $T(x_1, x_2) = (-x_2, x_1)$. Find SoT and ToS. Is S (or T) invertible?

Now, let us look at some examples involving the composite of linear operators.

Example 4: Let $T: \mathbf{R}^2 \to \mathbf{R}^3$ and $S: \mathbf{R}^3 \to \mathbf{R}^2$ be defined by

$T(x_1, x_2) = (x_1, x_2, x_1 + x_2)$ and $S(x_1, x_2, x_3) = (x_1, x_2)$. Find SoT and ToS.

Solution: First, note that $T \in L(\mathbf{R}^2, \mathbf{R}^3)$ and $S \in L(\mathbf{R}^3, \mathbf{R}^2)$. $\therefore$ SoT and ToS are both well defined linear operators. Now,

$S \circ T(x_1, x_2) = S(T(x_1, x_2)) = S(x_1, x_2, x_1 + x_2) = (x_1, x_2)$.

Hence, SoT = the identity transformation of $\mathbf{R}^2 = I_{\mathbf{R}^2}$.

Now,

$T \circ S(x_1, x_2, x_3) = T(S(x_1, x_2, x_3)) = T(x_1, x_2) = (x_1, x_2, x_1 + x_2)$.

In this case $S \circ T \in A(\mathbf{R}^2)$, while $T \circ S \in A(\mathbf{R}^3)$. Clearly, $S \circ T \neq T \circ S$.

Also, note that $S \circ T = I$, but $T \circ S \neq I$.

Remark: Even if SoT and ToS both being to A (V). SoT may not be equal to ToS. We give such an example below.

Example 5: Let $S, T \in A(\mathbf{R}^2)$ be defined by $T(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$ and $S(x_1, x_2) = (0, x_2)$. Show that $S \circ T \neq T \circ S$.

Solution: You can check that $S \circ T(x_1, x_2) = (0, x_1 - x_2)$ and $T \circ S(x_1, x_2) = (x_2, -x_2)$. Thus, $\exists (x_1, x_2) \in \mathbf{R}^2$ such that $S \circ T(x_1, x_2) \neq T \circ S(x_1, x_2)$ (for instance, $S \circ T(1, 1) \neq T \circ S(1,1)$). That is, $S \circ T \neq T \circ S$.

Note: Before checking whether SoT is a well defined linear operator. You must be sure that both S and T are well defined linear operators.

Now try to solve the following exercises.

E10) Let $T(x_1, x_2) = (0, x_1, x_2)$ and $S(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$. Find SoT and ToS. When is SoT = ToS?

E11) Let $T(x_1, x_2) = (2x_1, x_1 + 2x_2)$ for $(x_1, x_2) \in \mathbf{R}^2$, and $S(x_1, x_2, x_3) = (x_1 + 3x_2, 3x_1 - x_2, x_3)$ for $(x_1, x_2, x_3) \in \mathbf{R}^3$. Are SoT and ToS defined? If yes, find them.

E12) Let U, V, W, Z be vector spaces over F. Suppose $T \in L(U, V)$, $S \in L(V, W)$ and $R \in L(W, Z)$. Show that $(R \circ S) \circ T = R \circ (S \circ T)$.

E13) Let S, T ∈ A (V) and S be invertible. Show that rank (ST) = rank (TS) = rank (T). (ST means $S \circ T$.)

So far we have discussed the composition of linear transformations. We have seen that if S, T ∈ A (V), then $S \circ T$ ∈ A (V), where V is a vector space of dimension n. Thus, we have introduced another binary operation (see Sec. 1.5.2) in A (V), namely, the composition of operators, denoted by $\circ$. Remember, we already have the binary operations given in Sec. 6.2. In the following theorem we state some simple properties that involve all these operations.

Theorem 6: Let R, S, T ∈ A (V) and let $\alpha \in F$. Then

a) $R \circ (S+T) = R \circ S + R \circ T$, and
$(S+T) \circ R = S \circ R + T \circ R$.

b) $\alpha (S \circ T) = \alpha S \circ T = S \circ \alpha T$.

Proof: a) For any v ∈ V,

$$\begin{aligned} R \circ (S+T)(v) &= R((S+T)(v)) = R(S(v) + T(v)) \\ &= R(S(v)) + R(T(v)) \\ &= (R \circ S)(v) + (R \circ T)(v) \\ &= (R \circ S + R \circ T)(v) \end{aligned}$$

Hence, $R \circ (S+T) = R \circ S + R \circ T$.
Similarly, we can prove that $(S+T) \circ R = S \circ R + T \circ R$

b) For any v ∈ V,

$$\begin{aligned} \alpha (S \circ T)(v) &= \alpha (S(T(v))) \\ &= (\alpha S)(T(v)) \\ &= (\alpha S \circ T)(v) \end{aligned}$$

Therefore, $\alpha (S \circ T) = \alpha S \circ T$.

Similarly, we can show that $\alpha (S \circ T) = S \circ \alpha T$.

Notation: In future we shall be writing ST in place of $S \circ T$. Thus, ST (u) = S(T(u)) = $(S \circ T)$ u

Also, if T ∈ A (V), we write $T^0 = I$, $T^1 = T$, $T^2 = T \circ T$ and, in general, $T^n = T^{n-1} \circ T = T \circ T^{n-1}$.

The properties of A(V) stated in Theorems 1 and 6 are very important and will be used implicitly again and again. To get used to A(V) and the operations in it try the following exercises.

E14) Consider S, T: $\mathbf{R}^2 \to \mathbf{R}^2$ defined by S $(x_1, x_2) = (x_1, -x_2)$ and T $(x_1, x_2) = (x_1 + x_2, x_2 - x_1)$. What are S + T, ST, TS, $S \circ (S-T)$ and $(S-T) \circ S$?

E15) Let $S \in A(V)$, dim $V = n$ and rank $(S) = r$. Let
$M = \{T \in A(V) \mid ST = \mathbf{0}\}$,
$N = \{T \in A(V) \mid TS = \mathbf{0}\}$,

a) Show that M and N are subspaces of A (V).

b) Show that $M = L(V, \text{Ker } S)$. What is dim M?

By now you must have got used to handling the elements of A (V). The next section deals with polynomials that are related to these elements.

6.5 MINIMAL POLYNOMIAL

Recall that a polynomial in one variable x over F is of the form $p(x) = a_0 + a_1x + ... + a_n x^n$, where $a_0, a_1, ..., a_n \in F$.

If $a_n \neq 0$, then p (x) is said to be of degree n. If $a_n = 1$, then p (x) is called a monic polynomial of degree n. For example, $x^2 + 5x + 6$ is a monic polynomial of degree 2. The set of all polynomials in x with coefficients in F is denoted by F [x].

Definition: For a polynomial p, as above, and an operator $T \in A(V)$, we define $p(T) = a_0 I + a_1 T + ... + a_n T^n$

Since each of $I, T, ..., T^n \in A(V)$, we find $p(T) \in A(V)$. We say $p(T) \in F[T]$.

If q is another polynomial in x over F, then p (T) q (T) = q (T) p (T), that is, p (T) and q (T) commute with each other. This can be seen as follows:

Let $q(T) = b_0 I + b_1 T + ... + b_m T^m$

Then $p(T)\,q(T) = (a_0 I + a_1 T + ... + a_n T^n)(b_0 I + b_1 T + ... + b_m T^m)$

$$= a_0 b_0 I + (a_0 b_1 + a_1 b_0) T + ... + a_n b_m T^{n+m}$$

$$= (b_0 I + b_1 T + ... + b_m T^m)(a_0 I + a_1 T + \quad + a_n T^n)$$

$$= q(T)\,p(T).$$

E16) Let $p, q \in F[x]$ such that $p(T) = 0$, $q(T) = 0$. Show that $(p + q)(T) = 0$. ($(p + q)(x)$ means $p(x) + q(x)$.)

E17) Check that $(2I + 3S + S^3)$ commutes with $(S + 2S^4)$, for $S \in A(R^n)$.

We now go on to prove that given any $T \in A(V)$ we can find a polynomial $g \in F[x]$ such that

$g(T) = 0$, that is, $g(T)(v) = 0 \ \forall \ v \in V$.

Theorem 7: Let V be a vector space over **F** of dimension n and $T \in A(V)$. Then there exists a non-zero polynomial g over **F** such that $g(T) = 0$ and the degree of g is at most n^2.

Proof: We have already seen that $A(V)$ is a vector space of dimension n^2. Hence, the set $\{I, T, T^2, ..., T^{n^2}\}$ of $n^2 + 1$ vectors of $A(V)$, must be linearly dependent (ref. Unit 4, Theorem 7). Therefore, there must exist $a_0, a_1, ..., a_{n^2} \in F$ (not all zero) such that

$a_0 I + a_1 T + ... + a_{n^2} T^{n^2} = 0$.

Let g be the polynomial given by

$g(x) = a_0 + a_1 x + ... + a_{n^2} x^{n^2}$

Then g is a polynomial of degree at most n^2, such that $g(T) = 0$.

The following exercises will help you in getting used to polynomials in x and T.

E 18) Give an example of polynomials $g(x)$ and $h(x)$ in $R[x]$, for which $g(I) = 0$ and $h(0) = 0$, where I and 0 are the identity and zero transformations in $A(R^3)$.

E 19) Let $T \in A(V)$. Then we have a map ϕ from $F[x]$ to $A(V)$ given by $\phi(p) = p(T)$. Show that, for $a, b \in F$ and $p, q \in F[x]$,

a) $\phi(ap + bq) = a\phi(p) + b\phi(q)$,

b) $\phi(pq) = \phi(p)\phi(q)$.

*deg f denotes degree of the polynomial f.

In Theorem 7 we have proved that there exists some $g \in F[x]$ with $g(T) = 0$. But, if $g(T) = 0$, then $(\alpha g)(T) = 0$, for any $\alpha \in F$. Also, if $\deg g \le n^2$, then $\deg(\alpha g) \le n^2$. Thus, there are infinitely many polynomials that satisfy the conditions in Theorem 7. But if we insist on some more conditions on the polynomial g, then we end up with one and only one polynomial which will satisfy these conditions and the conditions in Theorem 7. Let us see what the conditions are.

Theorem 8: Let $T \in A(V)$. Then there exists a unique **monic polynomial p of smallest** degree such that $p(T) = 0$.

Proof: Consider the set $S = \{g \in F[x] \mid g(T) = 0\}$. This set is non-empty since, by Theorem 7, there exists a non-zero polynomial g, of degree at most n^2, such that $g(T) = 0$. Now consider the set $D = \{\deg f \mid f \in S\}$. Then D is a subset of $N \cup \{0\}$, and therefore, it must have a minimum element, say, m. Let $h \in S$ such that deg h = m. Then, h (T) = 0 and deg h ≤ deg g $\forall g \in S$.

If $h = a_0 + a_1 x + ... + a_m x^m$, $a_m \neq 0$, then $p = a_m^{-1} h$ is a monic polynomial such that p(T) = 0. Also deg p = deg h ≤ deg g $\forall g \in S$. Thus, we have shown that there exists a monic polynomial p, of least degree, such that p(T) = 0.

We now show that p is unique, that is, if q is any monic polynomial of smallest degree such that q (T) = 0, then p = q. But this is easy. Firstly, since deg p ≤ deg g $\forall g \in S$, deg p ≤ deg q. Similarly, deg q ≤ deg p. ∴ deg p = deg q.

Now suppose $p(x) = a_0 + a_1 x + ... + a_{n-1}x^{n-1} + x^n$ and $q(x) = b_0 + b_1 x + ... + b_{n-1} x^{n-1} + x^n$.

Since p (T) = 0 and q (T) = 0, we get (p − q) (T) = 0. But $p - q = (a_0 - b_0) + ... + (a_{n-1} - b_{n-1})x^{n-1}$. Hence, (p − q) is a polynomial of degree strictly less than the degree of p, such that (p − q) (T) = 0. That is, $p - q \in S$ with deg (p − q) < deg p. This is a contradiction to the way we chose p, unless p − q = 0, that is, p = q. ∴ p is the unique polynomial satisfying the conditions of Theorem 8.

This theorem immediately leads us to the following definition.

Definition: For $T \in A(V)$, the unique monic polynomial p of smallest degree such that p(T) = 0 is called the **minimal polynomial of T.**

Note that the minimal polynomial p, of T, is uniquely determined by the following three properties.

1) p is a monic polynomial over **F**.

2) p (T) = 0.

3) If $g \in F(x)$ with g (T) = 0, then deg p ≤ deg g.

Consider the following example and exercises.

Example 6: For any vector space V, find the minimal polynomials for I, the identity transformation, and 0, the zero transformation.

Solution: Let p (x) = x − 1 and q (x) = x. Then p and q are monic such that p(I) = 0 and q (0) = 0. Clearly no non-zero polynomials of smaller degree have the above properties. Thus, x − 1 and x are the required polynomials.

E20) Define $T: \mathbf{R}^3 \rightarrow \mathbf{R}^3$: $T(x_1, x_2, x_3) = (0, x_1, x_2)$. Show that the minimal polynomial of T is x^3.

E21) Define $T: \mathbf{R}^n \rightarrow \mathbf{R}^n$: $T(x_1, ..., x_n) = (0, x_1, ..., x_{n-1})$. What is the minimal polynomial of T? (Does E 20 help you?)

E22) Let $T:R^3 \to R^3$ be defined by
$T(x_1, x_2, x_3) = (3x_1, x_1 - x_2, 2x_1 + x_2 + x_3)$. Show that $(T^2 - I).(T - 3I) = 0$. What is the minimal polynomial of T?

We will now state and prove a criterion by which we can obtain the minimal polynomial of a linear operator T, once we know any polynomial $f \in F[x]$ with $f(T) = 0$. It says that the minimal polynomial must be a factor of any such f.

Theorem 9: Let $T \in A(V)$ and let p(x) be the minimal polynomial of T. Let f (x) be any polynomial such that f (T) = 0. Then there exists a polynomial g(x) such that f(x) = p(x)g(x).

Proof: The division algorithm states that given f (x) and p(x), there exist polynomials g (x) and h (x) such that f (x) = p (x) g (x) + h (x), where h (x) = 0 or deg h (x) < deg p (x). Now,

$0 = f(T) = p(T)\, g(T) + h(T) = h(T)$, since $p(T) = 0$.

Therefore, if $h(x) \neq 0$, then $h(T) = 0$, and deg h (x) < deg p (x).

This contradicts the fact that p (x) is the minimal polynomial of T. Hence, h (x) = 0, and we get f (x) = p (x) g (x).

Using this theorem, can you obtain the minimal polynomial of T in E22 more easily? Now we only need to check if $T - I$, $T + I$ or $T - 3I$ are 0.

Remark: If dim $V = n$ and $T \in A(V)$, we have seen that the degree of the minimal polynomial p of $T \le n^2$. In Unit 11, we shall see that the degree of p cannot exceed n. We shall also study a systematic method of finding the minimal polynomial of T, and some applications of this polynomial. But now we will only illustrate one application of the concept of the minimal polynomial by proving the following theorem.

Theorem 10: Let $T \in A(V)$. Then T is invertible if and only if the constant term in the minimal polynomial of T is not zero.

Proof: Let $p(x) = a_0 + a_1x + ... + a_m x^{m-1} + x^m$ be the minimal polynomial of T. Then $a_0 I + a_1 T + ... + a_{m-1} T^{m-1} + T^m = 0$.

$\Rightarrow T(a_1 I + ... + a_{m-1} T^{m-2} + T^{m-1}) = -a_0 I$ .... (1)

Firstly, we will show that if T^{-1} exists, then $a_0 \neq 0$. On the contrary, suppose $a_0 = 0$. Then (1) implies that $T(a_1 I + ... + T^{m-1}) = 0$. Multiplying both sides by T^{-1} on the left, we get $a_1 I + ... + T^{m-1} = 0$.

This equation gives us a monic polynomial $q(x) = a_1 + ... + x^{m-1}$ such that $q(T) = 0$ and deg q < deg p. This contradicts the fact that p is the minimal polynomial of T. Therefore, if T^{-1} exists then the constant term in the minimal polynomial of T cannot be zero.

Conversely, suppose the constant term in the minimal polynomial of T is not zero, that is, $a_0 \neq 0$. Then dividing Equation (1) on both sides by $(-a_0)$, we get

$T((-a_1/a_0) I + ... + (-1/a_0) T^{m-1}) = I$.

Let $S = (-a_1/a_0) I + ... + (-1/a_0) T^{m-1}$

Then we have $ST = I$ and $TS = I$. This shows, by Theorem 5, that T^{-1} exists and $T^{-1} = S$.

E23) Let P_n be the space of all polynomials of degree $\le n$. Consider the linear operator $D: P_2 \to P_2$ given by $D(a_0 + a_1 x + a_2 x^2) = a_1 + 2a_2 x$. (Note that D is just the differentiation operator.) Show that $D^3 = 0$. What is the minimal polynomial of D? Is D invertible?

E24) Consider the reflection transformation given in Unit 5, Example 4. Find its minimal polynomial. Is T invertible? If so, find its inverse.

E25) Let the minimal polynomial of $S \in A(V)$ be x^n, $n \ge 1$. Show that there exists $v_0 \in V$ such that the set $\{v_0, S(v_0), ..., S^{n-1}(v_0)\}$ is linearly independent.

We will now end the unit by summarising what we have covered in it.

6.6 SUMMARY

In this unit we covered the following points.

1) L (U, V), the vector space of all linear transformations from U to V is of dimension (dim U) (dim V).

2) The dual space of a vector space V is L (V, F)=V^*, and is isomorphic to V.

3) If $\{e_1, \ldots, e_n\}$ is a basis of V and $\{f_1, \ldots, f_n\}$ is its dual basis, then $f = \sum_{i=1}^{n} f(e_i) f_i \; \forall \; f \in V^*$ and $v = \sum_{i=1}^{n} f_i(v) e_i \; \forall \; v \in V$

4) Every vector space is isomorphic to its second dual.

5) Suppose S ∈ L (V, W) and T ∈ L (U, V). Then their composition $S \circ T$ ∈ L (U, W).

6) S ∈ A (V) = L(V, V) is an isomorphism if and only if there exists T ∈ A (V) such that $S \circ T = I = T \circ S$.

7) For T ∈ A (V) there exists a non-zero polynomial g ∈ F [x], of degree at most n^2, such that g (T) = 0, where dim V= n.

8) The minimal polynomial of T ∈ A (V) is the monic polynomial p, of smallest degree, such that p (T) = 0.

9) If p is the minimal polynomial of T and f is a polynomial such that f (T) = 0, then there exists a polynomial g (x) such that f (x) = p (x) g (x).

10) Let T ∈ A (V). Then T^{-1} exists if and only if the constant term in the minimal polynomial of T is not zero.

6.7 SOLUTIONS/ANSWERS

E1) We have to check that VS1—VS10 are satisfied by L(U, V). We have already shown that VS1 and VS6 are true.

VS2: For any L, M, N ∈ L (U, V), we have ∀u ∈ U, [(L+M) + N] (u)

= (L+M) (u) + N (u) = [L (u) + M (u)] + N (u)

= L (u) + [M(u) + N (u)], since addition is associative in V.

= [L + (M + N)] (u)

∴ (L + M) + N = L + (M + N).

VS3: 0: U→V: 0 (u) = 0 ∀u ∈ U is the zero element of L (U, V).

VS4: For any $S \in L(U, V)$, $(-1)S = -S$, is the additive inverse of S.

VS5: Since addition is commutative in V, $S + T = T + S \ \forall S, T$ in $L(U, V)$.

VS7: $\forall \alpha \in \mathbf{F}$ and $S, T \in L(U, V)$,

$\alpha(S + T)(u) = (\alpha S + \alpha T)(u) \ \forall u \in U$.

$\therefore \alpha(S + T) = \alpha S + \alpha T$.

VS8: $\forall \alpha, \beta \in \mathbf{F}$ and $S \in L(U, V)$, $(\alpha + \beta)S = \alpha S + \beta S$.

VS9: $\forall \alpha, \beta \in \mathbf{F}$ and $S \in L(U, V)$, $(\alpha\beta)S = \alpha(\beta S)$.

VS10: $\forall S \in L(U, V)$, $1.S = S$.

E2) $E_{2m}(e_m) = f_2$ and $E_{2m}(e_i) = 0$ for $i \neq m$

$E_{32}(e_2) = f_3$ and $E_{32}(e_i) = 0$ for $i \neq 2$.

$E_{mn}(e_i) = \begin{cases} f_m, & \text{if } i = n \\ 0 & \text{otherwise} \end{cases}$

E3) Both spaces have dimension 2 over R. A basis for $L(\mathbf{R}^2, \mathbf{R})$ is $\{E_{11}, E_{12}\}$, where $E_{11}(1, 0) = 1$, $E_{11}(0, 1) = 0$, $E_{12}(1, 0) = 0$, $E_{12}(0, 1) = 1$. A basis for $L(\mathbf{R}, \mathbf{R}^2)$ is $\{E_{11}, E_{21}\}$, where $E_{11}(1) = (1, 0)$, $E_{21}(1) = (0, 1)$.

E4) Let $f: \mathbf{R}^3 \to \mathbf{R}$ be any linear functional. Let $f(1, 0, 0) = a_1$, $f(0, 1, 0) = a_2$, $f(0, 0, 1) = a_3$. Then, for any $x = (x_1, x_2, x_3)$, we have $x = x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1)$.

$\therefore f(x) = x_1 f(1, 0, 0) + x_2 f(0, 1, 0) + x_3 f(0, 0, 1)$

$= a_1 x_1 + a_2 x_2 + a_3 x_3$.

E5) Let the dual basis be $\{f_1, f_2, f_3\}$. Then, for any $v \in \mathbf{P}_2$, $v = f_1(v).1 + f_2(v).x + f_3(v).x^2$.

$\therefore$, if $v = a_0 + a_1 x + a_2 x^2$, then $f_1(v) = a_0$, $f_2(v) = a_1$, $f_3(v) = a_2$.

That is, $f_1(a_0 + a_1 x + a_2 x^2) = a_0$, $f_2(a_0 + a_1 x + a_2 x^2) = a_1$, $f_3(a_0 + a_1 x + a_2 x^2) = a_2$, for any $a_0 + a_1 x + a_2 x^2 \in \mathbf{P}_2$.

E6) Let $\{f_1, \ldots, f_n\}$ be a basis of V^*. Let its dual basis be $\{\theta_1, \ldots, \theta_n\}$. $\theta_i \in V^{**}$. Let $e_i \in V$ such that $\phi(e_i) = \theta_i$ (ref. Theorem 3) for $i = 1, \ldots, n$.

Then $\{e_1, \ldots, e_n\}$ is a basis of V, since ϕ^{-1} is an isomorphism and maps a basis to $\{e_1, \ldots, e_n\}$. Now $f_i(e_j) = \phi(e_j)(f_i) = \theta_j(f_i) = \delta_{ji}$, by definition of a dual basis.

$\therefore \{f_1, \ldots, f_n\}$ is the dual of $\{e_1, \ldots, e_n\}$.

E7) For any $S \in A(V)$ and for any $v \in V$,

$S \circ I(v) = S(v)$ and $I \circ S(v) = I(S(v)) = S(v)$.

$\therefore S \circ I = S = I \circ S$.

E8) $\forall S \in A(V)$ and $v \in V$,

$S \circ 0(v) = S(0) = 0$, and

$0 \circ S(v) = 0(S(v)) = 0$.

$\therefore S \circ 0 = 0 \circ S = 0$.

E9) $S \in A(\mathbf{R}^2)$, $T \in A(\mathbf{R}^2)$.

$S \circ T(x_1, x_2) = S(-x_2, x_1) = (x_1, x_2)$

$T \circ S(x_1, x_2) = T(x_2, -x_1) = (x_1, x_2)$

$\forall (x_1, x_2) \in \mathbf{R}^2$.

$\therefore S \circ T = T \circ S = I$, and hence, both S and T are invertible.

E10) $T \in L(\mathbf{R}^2, \mathbf{R}^3)$, $S \in L(\mathbf{R}^3, \mathbf{R}^2)$. $\therefore S \circ T \in A(\mathbf{R}^2)$, $T \circ S \in A(\mathbf{R}^3)$.

$\therefore S \circ T$ and $T \circ S$ can never be equal.

Now, $S \circ T(x_1, x_2) = S(0, x_1, x_2) = (x_1, x_1 + x_2) \ \forall (x_1, x_2) \in \mathbf{R}^2$.

Also, $T \circ S(x_1, x_2, x_3) = T(x_1 + x_2, x_2 + x_3) = (0, x_1 + x_2, x_2 + x_3) \ \forall (x_1, x_2, x_3) \in \mathbf{R}^3$.

E11) Since $T \in A(\mathbf{R}^2)$ and $S \in A(\mathbf{R}^3)$, $S \circ T$ and $T \circ S$ are not defined.

E12) Both $(R \circ S) \circ T$ and $R \circ (S \circ T)$ are in $L(U, Z)$. For any $u \in U$,

$[(R \circ S) \circ T](u) = (R \circ S)[T(u)] = R[S(T(u))] = R[(S \circ T)(u)] = [R \circ (S \circ T)](u)$.

$\therefore (R \circ S) \circ T = R \circ (S \circ T)$.

E13) By Unit 5, Theorem 6, rank $(S \circ T) \leq$ rank (T).
Also, rank (T) = rank $(I \circ T)$ = rank $((S^{-1} \circ S) \circ T)$
= rank $(S^{-1} \circ (S \circ T)) \leq$ rank $(S \circ T)$ (by Unit 5, Theorem 6).
Thus, rank $(S \circ T) \leq$ rank $(T) \leq$ rank $(S \circ T)$.
$\therefore$ rank $(S \circ T)$ = rank (T).
Similarly, you can show that rank $(T \circ S)$ = rank (T).

E14) $(S + T)(x, y) = (x, -y) + (x + y, y - x) = (2x + y, -x)$
$ST(x, y) = S(x + y, y - x) = (x + y, x - y)$
$TS(x, y) = T(x, -y) = (x - y, -(x + y))$
$[S \circ (S - T)](x, y) = S(-y, x - 2y) = (-y, 2y - x)$
$[(S - T) \circ S](x, y) = (S - T)(x, -y) = (x, y) - (x - y, -(x + y)) = (y, 2y + x)$
$\forall (x, y) \in \mathbf{R}^2$.

E15) a) We first show that if $A, B \in M$ and $\alpha, \beta \in F$, then $\alpha A + \beta B \in M$. Now,
$S \circ (\alpha A + \beta B) = S \circ \alpha A + S \circ \beta B$, by Theorem 6.
$= \alpha (S \circ A) + \beta (S \circ B)$, again by Theorem 6.
$= \alpha 0 + \beta 0$, since $A, B \in M$.
$= 0$
$\therefore \alpha A + \beta B \in M$, and M is a subspace of A (V).
Similarly, you can show that N is a subspace of A (V).

b) For any $T \in M$, $ST(v) = 0 \ \forall v \in V$. $\therefore T(v) \in \text{Ker } S \ \forall v \in V$.
$\therefore$ R (T), the range of T, is a subspace of Ker S.
$\therefore T \in L(V, \text{Ker } S)$. $\therefore M \subseteq L(V, \text{Ker } S)$.
Conversely, for any $T \in L(V, \text{Ker } S)$, $T \in A(V)$ such that $S(T(v)) = 0 \ \forall v \in V$.
$\therefore ST = 0$. $\therefore T \in M$.
$\therefore L(V, \text{Ker } S) \subseteq M$.
$\therefore$ We have proved that $M = L(V, \text{Ker } S)$.
$\therefore \dim M = (\dim V)(\text{nullity } S)$, by Theorem 1.
$= n(n - r)$, by the Rank Nullity Theorem.

E16) $(p+q)(T) = p(T) + q(T) = 0 + 0 = 0$.

E17) $(2I + 3S + S^3)(S + 2S^4) = (2I + 3S + S^3)S + (2I + 3S + S^3)(2S^4)$
$= 2S + 3S^2 + S^4 + 4S^4 + 6S^5 + 2S^7$
$= 2S + 3S^2 + 5S^4 + 6S^5 + 2S^7$
Also, $(S + 2S^4)(2I + 3S + S^3) = 2S + 3S^2 + 5S^4 + 6S^5 + 2S^7$
$\therefore$, $(S + 2S^4)(2I + 3S + S^3) = (2I + 3S + S^3)(S + 2S^4)$.

E18) Consider $g(x) = x - 1 \in \mathbf{R}[x]$. Then $g(I) = I - 1.I = 0$.
Also, if $h(x) = x$, then $h(0) = 0$.
Notice that the degrees of g and h are both $1 \leq \dim \mathbf{R}^3$.

E19) Let $p = a_0 + a_1 x + \dots + a_n x^n$, $q = b_0 + b_1 x + \dots + b_m x^m$.

a) Then $ap + bq = aa_0 + aa_1 x + \dots + aa_n x^n + bb_0 + bb_1 x + \dots + bb_m x^m$.
$\therefore \phi(ap + bq) = aa_0 I + aa_1 T + \dots + aa_n T^n + bb_0 I + bb_1 T + \dots + bb_m T^m$
$= ap(T) + bq(T) = a\phi(p) + b\phi(q)$

b) $pq = (a_0 + a_1 x + \dots + a_n x^n)(b_0 + b_1 x + \dots + b_m x^m)$
$= a_0 b_0 + (a_1 b_0 + a_0 b_1) x + \dots + a_n b_m x^{n+m}$
$\therefore \phi(pq) = a_0 b_0 I + (a_1 b_0 + a_0 b_1) T + \dots + a_n b_m T^{n+m}$
$= (a_0 I + a_1 T + \dots + a_n T^n)(b_0 I + b_1 T + \dots + b_m T^m)$
$= \phi(p) \circ \phi(q)$.

E20) $T \in A(\mathbf{R}^3)$. Let $p(x) = x^3$. Then p is a monic polynomial. Also, $p(T)(x_1, x_2, x_3) =$
$T^3(x_1, x_2, x_3) = T^2(0, x_1, x_2) = T(0, 0, x_1) = (0, 0, 0) \ \forall (x_1, x_2, x_3) \in \mathbf{R}^3$.
$\therefore p(T) = 0$.

We must also show that no monic polynomial q of smaller degree exists such that $q(T) = 0$.

Suppose $q = a+bx+x^2$ and $q(T) = 0$.

Then $(aI + bT + T^2)(x_1, x_2, x_3) = (0, 0, 0)$

$\Leftrightarrow a(x_1, x_2, x_3) + b(0, x_1, x_2) + (0, 0, x_1) = (0, 0, 0)$

$\Leftrightarrow ax_1 = 0, ax_2 + bx_1 = 0, ax_3 + bx_2 + x_1 = 0 \forall (x_1, x_2, x_3) \in R^3$.

$\Leftrightarrow a = 0, b = 0$ and $x_1 = 0$. But x_1 can be non-zero.

$\therefore$ q does not exist.

$\therefore$ p is a minimal polynomial of T.

E21) Consider $p(x) = x^n$. Then $p(T) = 0$ and no non-zero polynomial q of lesser degree exists such that $q(T) = 0$. This can be checked on the lines of the solution of E20.

E22) $(T^2-I)(T-3I)(x_1, x_2, x_3)$
$= (T^2-I)((3x_1, x_1-x_2, 2x_1+x_2+x_3) - (3x_1, 3x_2, 3x_3))$
$= (T^2-I)(0, x_1-4x_2, 2x_1+x_2-2x_3)$
$= T(0, -x_1+4x_2, 3x_1-3x_2-2x_3) - (0, x_1-4x_2, 2x_1+x_2-2x_3)$
$= (0, x_1-4x_2, 2x_1+x_2-2x_3) - (0, x_1-4x_2, 2x_1+x_2-2x_3)$
$= (0, 0, 0) \forall (x_1, x_2, x_3) \in R^3$.
$\therefore (T^2-I)(T-3I) = 0$

Suppose $\exists\, q = a+bx+x^2$ such that $q(T) = 0$. Then $q(T)(x_1, x_2, x_3) = (0, 0, 0) \forall (x_1, x_2, x_3) \in R^3$. This means that $a+3b+9 = 0$, $(b+2)x_1 + (a-b+1)x_2 = 0$, $(2b+9)x_1 + bx_2 + (a+b+1)x_3 = 0$. Eliminating a and b, we find that these equations can be solved provided $5x_1 - 2x_2 - 4x_3 = 0$. But they should be true for any $(x_1, x_2, x_3) \in R^3$. $\therefore$, the equations can't be solved, and q does not exist. $\therefore$, the minimal polynomial of T is $(x^2-1)(x-3)$.

E23) $D^4(a_0 + a_1x + a_2x^2) = D^3(a_1 + 2a_2x) = D^2(2a_2) = D(0) = 0 \forall a_0 + a_1x + a_2x^2 \in P_2$.
$\therefore D^4 = 0$.

The minimal polynomial of D can be D, D^2, D^3 or D^4. Check that $D^3 = 0$, but $D^2 \neq 0$. $\therefore$ the minimal polynomial of D is $p(x) = x^3$. Since p has no non-zero constant term, D is not an isomorphism.

E24) $T: R^2 \to R^2: T(x,y) = (x, -y)$.
Check that $T^2 - I = 0$
$\therefore$ the minimal polynomial p must divide x^2-1.
$\therefore p(x)$ can be $x-1$, $x+1$ or x^2-1. Since $T-I \neq 0$ and $T+I \neq 0$, we see that $p(x) = x^2-1$.
By Theorem 10, T is invertible. Now $T^2 - I = 0$.
$\therefore T(-T) = I$. $\therefore T^{-1} = -T$.

E25) Since the minimal polynomial of S is x^n, $S^n = 0$ and $S^{n-1} \neq 0$. $\therefore \exists\, v_0 \in V$ such that $S^{n-1}(v_0) \neq 0$. Let $a_1, a_2, \ldots, a_n \in F$ such that
$a_1 v_0 + a_2 S(v_0) + \ldots + a_n S^{n-1}(v_0) = 0$. ...(1)

Then, applying S^{n-1} to both sides of this equation, we get $a_1 S^{n-1}(v_0) + a_2 S^n(v_0) + \ldots + a_n S^{2n-1}(v_0) = 0$.
$\Rightarrow a_1 S^{n-1}(v_0) = 0$, since $S^n = 0 = S^{n+1} = \ldots = S^{2n-1}$
$\Rightarrow a_1 = 0$.
Now (1) reduces to $a_2 S(v_0) + \ldots + a_n S^{n-1}(v_0) = 0$.
Applying S^{n-2} to both sides we get $a_2 = 0$. In this way we get $a_i = 0 \forall i = 1, \ldots, n$.
$\therefore$ The set $\{v_0, S(v_0), \ldots, S^{n-1}(v_0)\}$ is linearly independent.

UNIT 7 MATRICES - I

Structure

7.1 INTRODUCTION

You have studied linear transformations in Units 5 and 6. We will now study a simple means of representing them, namely, by matrices (the plural form of 'matrix'). We will show that, given a linear transformation, we can obtain a matrix associated to it, and vice versa. Then, as you will see, certain properties of a linear transformation can be studied more easily if we study the associated matrix instead. For example, you will see in Block 3, that it is often easier to obtain the characteristic roots of a matrix than of a linear transformation.

Matrices were introduced by the English mathematician, Arthur Cayley, in 1858. He came upon this notion in connection with linear substitutions. Matrix theory now occupies an important position in pure as well as applied mathematics. In physics one comes across such terms as matrix mechanics, scattering matrix, spin matrix, annihilation and creation matrices. In economics we have the input-output matrix and the pay off matrix; in statistics we have the transition matrix; and, in engineering, the stress matrix, strain matrix, and many other matrices.

Matrices are intimately connected with linear transformations. In this unit we will bring out this link. We will first define matrices and derive algebraic operations on matrices from the corresponding operations on linear transformations. We will also discuss some special types of matrices. One type, a triangular matrix, will be used often in Unit 8. You will also study invertible matrices in some detail, and their connection with change of bases. In Block 3 we will often refer to the material on change of bases, so do spend some time on Sec. 7.6.

To realise the deep connection between matrices and linear transformations, you should go back to the exact spot in Units 5 and 6 to which frequent references are made.

This unit may take you a little longer to study, than previous ones, but don't let that worry you. The material in it is actually very simple.

Objectives

After studying this unit, you should be able to

- define and give examples of various types of matrices;
- obtain a matrix associated to a given linear transformation;

- define a linear transformation, if you know its associated matrix;
- evaluate the sum, difference, product and scalar multiples of matrices;
- obtain the transpose and conjugate of a matrix;
- determine if a given matrix is invertible;
- obtain the inverse of a matrix;
- discuss the effect that the change of basis has on the matrix of a linear transformation.

7.2 VECTOR SPACE OF MATRICES

Consider the following system of three simultaneous equations in four unknowns:

$$\begin{aligned} x - 2y + 4z + t &= 0 \\ x + \tfrac{1}{2}y \qquad + 11t &= 0 \\ 3y - 5z \qquad &= 0 \end{aligned}$$

The coefficients of the unknowns, x, y, z and t, can be arranged in rows and columns to form a rectangular array as follows:

1	−2	4	1	(coefficients of the first equation)
1	1/2	0	11	(coefficients of the second equation)
0	3	−5	0	(coefficients of the third equation)

Such a rectangular array (or arrangement) of numbers is called a matrix. A matrix is usually enclosed within square brackets [ ] or round brackets ( ) as

$$\begin{bmatrix} 1 & -2 & 4 & 1 \\ 1 & \frac{1}{2} & 0 & 11 \\ 0 & 3 & -5 & 0 \end{bmatrix} \text{ or } \begin{pmatrix} 1 & -2 & 4 & 1 \\ 1 & \frac{1}{2} & 0 & 11 \\ 0 & 3 & -5 & 0 \end{pmatrix}$$

The numbers appearing in the various positions of a matrix are called the **entries** (or **elements**) **of the matrix**. Note that the same number may appear at two or more different positions of a matrix. For example, 1 appears in 3 different positions in the matrix given above.

In the matrix above, the three horizontal rows of entries have 4 elements each. These are called the **rows** of this matrix. The four vertical rows of entries in the matrix, having 3 elements each, are called its **columns**. Thus, this matrix has three rows and four columns. We describe this by saying that this is a matrix of size 3 × 4 ("3 by 4" or "3 cross 4"), or that this is a 3 × 4 matrix. The rows are counted from top to bottom and the columns are counted from left to right. Thus, the first row is (1, −2, 4, 1), the second row is (1, $\frac{1}{2}$, 0, 11), and so on. Similarly,

the first column is $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, the second column is $\begin{bmatrix} -2 \\ \frac{1}{2} \\ 3 \end{bmatrix}$, and so on.

Note that each row is a 1 × 4 matrix and each column is a 3 × 1 matrix.

We will now define a matrix of any size.

7.2.1 Definition of a Matrix

Let us see what we mean by a matrix of size m × n, where m and n are any two natural numbers.

Let F be a field.

A rectangular array

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

of mn elements of F arranged in m rows and n columns is called a **matrix of size m × n**, or an m × n matrix, over F. You must remember that the mn entries need not be distinct.

The element at the intersection of the ith row and the jth column is called the (i, j) th element. For example, in the m × n matrix above, the (2, n) th element is a_{2n}, which is the intersection of the 2nd row and the nth column.

A brief notation for this matrix is $[a_{ij}]_{m \times n}$, or simply $[a_{ij}]$, if m and n need not be stressed. We also denote matrices by capital letters A, B, C, ..., etc. The set of all m × n matrices over F is denoted by $M_{m \times n}(F)$.

Thus, $[1, \sqrt{2}\,] \in M_{1 \times 2}(\mathbf{R})$.

If m = n, then the matrix is called a **square matrix**. The set of all n × n matrices over F is denoted by $M_n(F)$.

In an m × n matrix each row is a 1 × n matrix and is also called a **row vector**. Similarly, each column is an m × 1 matrix and is also called a **column vector**.

Let us look at a situation in which a matrix can arise.

Example 1: There are 20 male and 5 female students in the B.Sc. (Math. Hon's) I year class in a certain college, 15 male and 10 female students in B.Sc. (Math. Hon's) II year and 12 male and 10 female students in B.Sc. (Math. Hon's) III year. How does this information give rise to a matrix?

Solution: One of the ways in which we can arrange this information in the form of a matrix is as follows:

	B.Sc. I	B.Sc. II	B.Sc. III
Male	20	15	12
Female	5	10	10

This is a 2 × 3 matrix.

Another way could be the 3 × 2 matrix

	Female	Male
B.Sc I	5	20
B.Sc. II	10	15
B.Sc III	10	12

Either of these matrix representations immediately shows us how many male/female students there are in any class.

To get used to matrices and their elements, you can try the following exercises

E 1) Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 0 \\ 0 & 0 & 7 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 5 & 3 & 2 \\ 5 & 4 & 1 & 5 \\ 0 & 3 & 2 & 0 \end{bmatrix}$

Give the

a) (1, 2)th elements of A and B.

b) third row of A.

c) second column of A and the first column of B.

d) fourth row of B.

E2) Write two different 4 × 2 matrices.

How did you solve E 2? Did the (i, j)th entry of one differ from the (i,j)th entry of the other for some i and j? If not, then they were equal. For example, the two 1×1 matrices [2] and [2] are equal. But $[2] \neq [3]$, since their entries at the (1, 1) position differ.

Definition: Two matrices are said to be **equal** if

i) they have the same size, that is, they have the same number of rows as well as the same number of columns, and

ii) their elements, at all the corresponding positions, are the same.

The following example will clarify what we mean by equal matrices.

Example 2: If $\begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} x & y \\ z & 3 \end{bmatrix}$, then what are x, y and z?

Solution: Firstly, both matrices are of the same size, namely, 2×2. Now, for these matrices to be equal the (i, j) th elements of both must be equal $\forall$ i, j. Therefore, we must have $x = 1$, $y = 0$, $z = 2$.

E E 3) Are [1] and $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ equal? Why?

Now that you are familiar with the concept of a matrix, we will link it up with linear transformations.

7.2.2 Matrix of a Linear Transformation

We will now obtain a matrix that corresponds to a given linear transformation. You will see how easy it is to go from matrices to linear transformations, and back.

Let U and V be vector spaces over a field F, of dimensions n and m, respectively. Let

$B_1 = \{e_1, \ldots\ldots, e_n\}$ be an ordered basis of U, and

$B_2 = \{f_1, \ldots\ldots, f_m\}$ be an ordered basis of V. (By an **ordered basis** we mean that the order in which the elements of the basis are written is fixed. Thus, an ordered basis $\{e_1, e_2\}$ is not equal to an ordered basis $\{e_2, e_1\}$.)

Given a linear transformation $T: U \to V$, we will associate a matrix to it. For this, we consider $T(e_1), \ldots\ldots, T(e_n)$, which are all elements of V and hence, they are linear combinations of $f_1, \ldots, f_m$. Thus, there exist mn scalars α_{ij}, such that

$T(e_1) = \alpha_{11} f_1 + \alpha_{21} f_2 + \ldots\ldots + \alpha_{m1} f_m$

....

$T(e_j) = \alpha_{1j} f_1 + \alpha_{2j} f_2 + \ldots\ldots + \alpha_{mj} f_m$

....

$T(e_n) = \alpha_{1n} f_1 + \alpha_{2n} f_2 + \ldots\ldots + \alpha_{mn} f_m$

From these n equations we form an $m \times n$ matrix whose first column consists of the coefficients of the first equation, second column consists of the coefficients of the second equation, and so on. This matrix,

$$A = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix}$$

is called the **matrix of T with respect to the bases B_1** and B_2. Notice that the **coordinate vector of $T(e_j)$ is the jth column of A.**

We use the notation $[T]_{B_1, B_2}$ for this matrix. Thus, to obtain $[T]_{B_1, B_2}$ we consider $T(e_i)\ \forall e_i \in B_1$, and write them as linear combinations of the elements of B_2.

If $T \in L(V, V)$, B is a basis of V and we take $B_1 = B_2 = B$, then $[T]_{B, B}$ is called the matrix of T with respect to the basis B, and can also be written as $[T]_B$.

Remark : Why do we insist on ordered bases? What happens if we interchange the order of the elements in B_1 to $\{e_n, e_1, ..., e_{n-1}\}$? The matrix $[T]_{B_1,B_2}$ also changes, the last column becoming the first column now. Similarly, if we change the positions of the f_i's in B_2, the rows of $[T]_{B_1,B_2}$ will get interchanged.

Thus, to obtain a unique matrix corresponding to T, we must insist on B_1 and B_2 being ordered bases. **Henceforth, while discussing the matrix of a linear mapping, we will always assume that our bases are ordered bases.**

We will now give an example, followed by some exercises.

Example 3: Consider the linear operator
$T : R^3 \to R^2 : T(x, y, z) = (x, y)$. Choose bases B_1 and B_2 of R^3 and R^2, respectively. Then obtain $[T]_{B_1,B_2}$.

Solution: Let $B_1 = \{e_1, e_2, e_3\}$, where $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. Let $B_2 = \{f_1, f_2\}$, where $f_1 = (1, 0)$, $f_2 = (0, 1)$. Note that B_1 and B_2 are the standard bases of R^3 and R^2, respectively.

$T(e_1) = (1, 0) = f_1 = 1.f_1 + 0.f_2$

$T(e_2) = (0, 1) = f_2 = 0.f_1 + 1.f_2$

$T(e_3) = (0, 0) = 0f_1 + 0f_2$.

Thus, $[T]_{B_1,B_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

E 4) Choose two other bases B_1' and B_2' of R^3 and R^2, respectively. (In Unit 4 you came across a lot of bases of both these vector spaces.) For T in the example above, give the matrix $[T]_{B_1',B_2'}$.

What E4 shows us is that the matrix of a transformation depends on the bases that we use for obtaining it. The next two exercises also bring out the same fact.

E 5) Write the matrix of the linear transformation $T: R^3 \to R^2 : T(x, y, z) = (x+2y+2z, 2x+3y+4z)$ with respect to the standard bases of R^3 and R^2.

E 6) What is the matrix of T, in E 5, with respect to the bases

$B_1' = \{(1, 0, 0), (0, 1, 0), (1, -2, 1)\}$ and

$B_2' = \{(1, 2), (2, 3)\}$?

The next exercise is about an operator that you have come across often.

E 7) Let V be the vector space of polynomials over R of degree ≤ 3, in the variable t. Let D: V → V be the differential operator given in Unit 5 (E6, when n = 3). Show that the matrix of D with respect to the basis {1, t, t^2, t^3} is

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

So far, given a linear transformation, we have obtained a matrix from it. This works the other way also. That is, given a matrix we can define a linear transformation corresponding to it.

Example 4 : Describe T:$\mathbf{R}^3 \rightarrow \mathbf{R}^3$ such that

$[T]_B = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$, where B is the standard basis of $\mathbf{R}^3$

Solution: Let B = {e_1, e_2, e_3}. Now, we are given that

$T(e_1) = 1.e_1 + 2.e_2 + 3.e_3$

$T(e_2) = 2.e_1 + 3.e_2 + 1.e_3$

$T(e_3) = 4.e_1 + 1.e_2 + 2.e_3$

You know that any element of $\mathbf{R}^3$ is $(x, y, z) = xe_1 + ye_2 + ze_3$.

Therefore, $T(x, y, z) = T(xe_1 + ye_2 + ze_3)$

$= xT(e_1) + yT(e_2) + zT(e_3)$, since T is linear.

$= x(e_1 + 2e_2 + 3e_3) + y(2e_1 + 3e_2 + e_3) + z(4e_1 + e_2 + 2e_3)$

$= (x+2y+4z)e_1 + (2x + 3y + z)e_2 + (3x + y + 2z)e_3$

$= (x+2y+4z, 2x+3y+z, 3x+y+2z)$

$\therefore$ T: $\mathbf{R}^3 \rightarrow \mathbf{R}^3$ is defined by $T(x, y, z) = (x + 2y + 4z, 2x + 3y + z, 3x + y + 2z)$

Try the following exercises now.

E 8) Describe T: $\mathbf{R}^3 \rightarrow \mathbf{R}^2$ such that

$[T]_{B_1, B_2} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, where B_1 and B_2 are the standard bases of $\mathbf{R}^3$ and $\mathbf{R}^2$, respectively.

E 9) Find the linear operator T: C → C whose matrix, with respect to the basis {1, i} is $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. (Note that C, the field of complex numbers, is a vector space over R, of dimension 2.)

Now we are in a position to define the sum of matrices and multiplication of a matrix by a scalar.

7.2.3 Sum and Multiplication by Scalars

In Unit 5 you studied about the sum and scalar multiples of linear transformations. In the following theorem we will see what happens to the matrices associated with the linear transformations that are sums or scalar multiples of given linear transformations.

Theorem 1: Let U and V be vector spaces over F, of dimensions n and m, respectively. Let B_1 and B_2 be arbitrary bases of U and V, respectively. (Let us abbreviate $[T]_{B_1, B_2}$ to [T] during this theorem.) Let S, T ∈ L(U, V) and α ∈ F. Suppose $[S] = [a_{ij}]$, $[T] = [b_{ij}]$. Then

$$[S + T] = [a_{ij} + b_{ij}], \text{ and}$$

$$[\alpha S] = [\alpha a_{ij}]$$

Proof: Suppose $B_1 = \{e_1, e_2, ..., e_n\}$ and $B_2 = \{f_1, f_2, ..., f_m\}$. Then all the matrices to be considered here will be of size m × n.

Now, by our hypothesis,

$$S(e_j) = \sum_{i=1}^{m} a_{ij} f_i \ \forall j = 1, \ldots, n \text{ and}$$

$$T(e_j) = \sum_{i=1}^{m} b_{ij} f_i \ \forall j = 1, \ldots, n$$

$\therefore$, $(S + T)(e_j) = S(e_j) + T(e_j)$ (by definition of S + T)

$$= \sum_{i=1}^{m} a_{ij} f_i + \sum_{i=1}^{m} b_{ij} f_i$$

$$= \sum_{i=1}^{m} (a_{ij} + b_{ij}) f_i$$

Thus, by definition of the matrix with respect to B_1 and B_2, we get $[S + T] = [a_{ij} + b_{ij}]$.

Now, $(\alpha S)(e_j) = \alpha(S(e_j))$ (by definition of αS)

$$= \alpha\left(\sum_{i=1}^{m} a_{ij} f_i\right)$$
$$= \sum_{i=1}^{m} (\alpha a_{ij}) f_i$$

Thus, $[\alpha S] = [\alpha a_{ij}]$

Theorem 1 motivates us to define the sum of 2 matrices in the following way

Two matrices can be added if and only if they are of the same size

Definition: Let A and B be the following two m × n matrices.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

Then the sum of A and B is defined to be the matrix

$$A + B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

In other words, A + B is the m × n matrix whose (i, j) th element is the sum of the (i, j) th element of A and the (i, j)th element of B.

Let us see an example of how two matrices are added.

Example 5: What is the sum of $\begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 4 & 5 \end{bmatrix}$?

Solution: Firstly, notice that both the matrices are of the same size (otherwise, we can't add them). Their sum is

$$\begin{bmatrix} 1+0 & 4+1 & 5+0 \\ 0+1 & 1+4 & 0+5 \end{bmatrix} = \begin{bmatrix} 1 & 5 & 5 \\ 1 & 5 & 5 \end{bmatrix}$$

E 10) What is the sum of

a) $[1 \quad 2]$ and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$?

b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$?

Now, let us define the scalar multiple of a matrix, again motivated by Theorem 1.

Definition: Let α be a scalar, i.e., $\alpha \in F$, and let $A = [a_{ij}]_{m \times n}$. Then we define the **scalar multiple of the matrix A by the scalar** α to be the matrix

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{bmatrix}$$

In other words, αA is the $m \times n$ matrix whose (i, j) th element is α times the (i, j) th element of A.

Example 6: What is 2A, where $A = \begin{bmatrix} 1/2 & 1/4 & 1/3 \\ 0 & 0 & 0 \end{bmatrix}$?

Solution: We must multiply each entry of A by 2 to get 2A.
Thus,

$$2A = \begin{bmatrix} 1 & 1/2 & 2/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

E E 11) Calculate $3\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $3\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $3\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$.

Remark: The way we have defined the sum and scalar multiple of matrices allows us to write Theorem 1 as follows:

$$[S + T]_{B_1, B_2} = [S]_{B_1, B_2} + [T]_{B_1, B_2}$$

$$[\alpha S]_{B_1, B_2} = \alpha [S]_{B_1, B_2}$$

The following exercise will help you in checking if you have understood the contents of Sections 7.2.2 and 7.2.3.

E E 12) Define $S: R^2 \to R^3 : S(x,y) = (x,0,y)$ and $T : R^2 \to R^3 : T(x,y) = (0,x,y)$. Let B_1 and B_2 be the standard bases for $\mathbf{R}^2$ and $\mathbf{R}^3$, respectively.

Then what are $[S]_{B_1, B_2}$, $[T]_{B_1, B_2}$, $[S + T]_{B_1, B_2}$, $[\alpha S]_{B_1, B_2}$, for any $\alpha \in \mathbf{R}$.

We now want to show that the set of all m × n matrices over F is actually a vector space over F.

7.2.4 $M_{m\times n}$ (F) is a Vector Space

After having defined the sum and scalar multiplication of matrices, we enumerate the properties of these operations. This will ultimately lead us to prove that the set of all m× n matrices over F is a vector space over F. Do keep the properties VS1-VS10 (of Unit 3) in mind.

For any $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}] \in M_{m\times n}(F)$ and $\alpha, \beta \in F$, we have

i) Matrix addition is associative:

$(A + B) + C = A + (B + C)$, since

$(a_{ij} + b_{ij} + c_{ij} = a_{ij} + (b_{ij} + c_{ij}) \forall i, j$, as they are elements of a field.

ii) Additive identity: The matrix of the zero transformation (see Unit 5), with respect to any basis, will have 0 as all its entries. This is called the **zero matrix**. Consider the zero matrix **0**, of size m × n. Then, for any $A \in M_{m\times n}(F)$,

$A + 0 = 0 + A = A$,

since $a_{ij} + 0 = 0 + a_{ij} = a_{ij} \forall i, j$.

Thus, **0** is the additive identity for $M_{m\times n}(F)$.

iii) Additive inverse: Given $A \in M_{m\times n}(F)$ we consider the matrix $(-1)A$. Then

$A + (-1)A = (-1)A + A = 0$

This is because the (i,j)th element of $(-1)A$ is $-a_{ij}$, and $a_{ij} + (-a_{ij}) = 0 = (-a_{ij}) + a_{ij} \forall i,j$.

Thus, $(-1)A$ is the **additive inverse** of A. We denote $(-1)A$ by $-A$.

iv) Matrix addition is commutative:

$A + B = B + A$

This is true because $a_{ij} + b_{ij} = b_{ij} + a_{ij} \forall i,j$.

v) $\alpha(A + B) = \alpha A + \alpha B$.

vi) $(\alpha + \beta)A = \alpha A + \beta A$

vii) $(\alpha\beta)A = \alpha(\beta A)$

viii) $1.A = A$

E 13) Write out the formal proofs of the properties (v) – (viii), given above.

These eight properties imply that $M_{m\times n}$ **(F) is a vector space over F.**

Now that we have shown that $M_{m\times n}(F)$ is a vector space over **F**, we know it must have a dimension.

7.2.5 Dimension of $M_{m\times n}$ (F) over F

What is the dimension of $M_{m\times n}(F)$ over F? To answer this question we prove the following theorem. But, before you go further, check whether you remember the definition of a vector space isomorphism (Unit 5).

Theorem 2: Let U and V be vector spaces over F of dimensions n and m, respectively. Let B_1 and B_2 be a pair of bases of U and V, respectively. The mapping $\phi : L(U,V) \to M_{m\times n}(F)$, given by $\phi(T) = [T]_{B_1, B_2}$ is a vector space isomorphism.

Proof: The fact that ϕ is a linear transformation follows from Theorem 1. We proceed to show that the map is also 1 – 1 and onto. For the rest of the proof we shall denote $[S]_{B_1, B_2}$ by $[S]$ only, and take $B_1 = \{e_1, \ldots, e_n\}$, $B_2 = \{f_1, f_2, \ldots, f_m\}$.

ϕ is 1 – 1: Suppose $S, T \in L(U,V)$ be such that $\phi(S) = \phi(T)$.

Then $[S] = [T]$. Therefore, $S(e_j) = T(e_j) \forall e_j \in B_1$.

Thus, by Unit 5 (Theorem 1), we have $S = T$.

ϕ is onto: If $A \in \mathbf{M}_{m \times n}(F)$ we want to construct $T \in L(U,V)$ such that $\phi(T) = A$. Suppose $A = [a_{ij}]$. Let $v_1, \ldots, v_n \in V$ such that

$$v_j = \sum_{i=1}^{m} a_{ij} f_i \text{ for } j = 1, \ldots, n.$$

Then, by Theorem 3 of Unit 5, there exists a linear transformation $T \in L(U,V)$ such that

$$T(e_j) = v_j = \sum_{i=1}^{n} a_{ij} f_i.$$

Thus, by definition, $\phi(T) = A$.

Therefore, ϕ is a vector space isomorphism.

A corollary to this theorem gives us the dimension of $\mathbf{M}_{m \times n}(F)$.

Corollary: Dimension of $\mathbf{M}_{m \times n}(F) = mn$.

Proof: Theorem 2 tells us that $\mathbf{M}_{m \times n}(F)$ is isomorphic to $L(U,V)$. Therefore, $\dim_F \mathbf{M}_{m \times n}(F) = \dim_F L(U,V)$ (by Theorem 12 of Unit 5) $= mn$, from Unit 6 (Theorem 1).

Why do you think we chose such a roundabout way for obtaining $\dim \mathbf{M}_{m \times n}(F)$? We could as well have tried to obtain mn linearly independent $m \times n$ matrices and show that they generate $\mathbf{M}_{m \times n}(F)$. But that would be quite tedious (see E16). Also, we have done so much work on $L(U,V)$ so why not use that! And, doesn't the way we have used seem neat?

Now for some exercises related to Theorem 2.

E E 14) At most, how many matrices can there be in any linearly independent subset of $\mathbf{M}_{2 \times 3}(F)$?

E E 15) Are the matrices [1, 0] and [1, -1] linearly independent over **R**?

E E 16) Let E_{ij} be an $m \times n$ matrix whose (i,j) th element is 1 and the other elements are 0. Show that $\{E_{ij} : 1 \le i \le m, 1 \le j \le n\}$ is a basis of $\mathbf{M}_{m \times n}(F)$ over **F**. Conclude that $\dim_F \mathbf{M}_{m \times n}(F) = mn$.

'ow we move on to the next section, where we see some ways of getting new matrices from iven ones.

.3 NEW MATRICES FROM OLD

iven any matrix we can obtain new matrices from them in different ways. Let us see three ' these ways.

3.1 Transpose

ippose $A = \begin{bmatrix} 1 & 0 & 9 \\ 2 & 5 & 9 \end{bmatrix}$

om this we form a matrix whose first and second columns are the first and second rows of respectively. That is, we obtain

$= \begin{bmatrix} 1 & 2 \\ 0 & 5 \\ 9 & 9 \end{bmatrix}$.

en B is called the transpose of A. Note that A is also the transpose of B, since the rows of ire the columns of A. Here A is a 2×3 matrix and B is a 3×2 matrix.

general, if $A = [a_{ij}]$ is an $m \times n$ matrix. Then the $n \times m$ matrix whose ith column is the ith v of A, is called the transpose of A. The transpose of A is denoted by A^t (The notation 1 A' is also widely used.)

te that, if $A = [a_{ij}]_{m \times n}$, then $A^t = [b_{ij}]_{n \times m}$ where b_{ij} is the intersection of the ith row and the column of A^t. $\therefore$ b_{ij} is the intersection of the jth row and ith column of A, i.e., a_{ji}.
$b_{ij} = a_{ji}$.

7) Find A^t, where $a = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$.

now give theorem that lists some properties of the transpose.

eorem 3: Let $A, B \in M_{m \times n}(F)$ and $\alpha \in F$. Then,

$(A + B)^t = A^t + B^t$

$(\alpha A)^t = \alpha A^t$.

$(A^t)^t = A$

Proof: a) Let $A = [a_{ij}]$ and $B = [b_{ij}]$. Then $A + B = [a_{ij} + b_{ij}]$.
Therefore, $(A + B)^t = [c_{ij}]$, where

c_{ij} = the (j,i)th element of $A+B = a_{ji} + b_{ji}$
= sum of the (j,i)th elements of A and B
= sum of the (i,j)th elements of A^t and B^t.
= (i,j)th element of $A^t + B^t$.

Thus, $(A + B)^t = A^t + B^t$

We leave you to complete the proof of this theorem. In fact that is what E 18 says!

E E 18) Prove (b) and (c) of Theorem 3.

E E 19) Show that, if $A = A^t$, then A must be a square matrix.

E 19 leads us to some definitions.

Definitions : A square matrix A such that $A^t = A$ is called a **symmetric matrix**. A square matrix A such that $A^t = -A$, is called a **skew-symmetric matrix**.
For example, the matrix in E 17, and

$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, are both symmetric matrices.

$\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$ is an example of a skew-symmetric matrix, since

$$\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}^t = \begin{bmatrix} 0 & -2 \\ 2 & 0 \end{bmatrix} = -\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}.$$

E E 20) Take a 2×2 matrix A. Calculate $A + A^t$ and $A - A^t$. Which of these is symmetric and which is skew-symmetric?

Every square matrix can be expressed as the sum of a symmetric and a skew-symmetric matrix.

What you have shown in E20 is true for a square matrix of any size, namely, for any $A \in M_n(F)$, $A + A^t$ is symmetric and $A - A^t$ is skew-symmetric.

We now give another way of getting a new matrix from a given matrix over the complex field.

7.3.2 Conjugate

If A is a matrix over C, then the matrix obtained by replacing each entry of A by its complex conjugate is called the **conjugate of A**, and is denoted by $\bar{A}$.

The complex conjugate of $a+ib \in C$ is $a-ib$.

Three properties of conjugates; which are similar to those of the transpose, are

a) $\overline{A+B} = \bar{A} + \bar{B}$, for $A, B \in M_{m\times n}(C)$

b) $\overline{\alpha A} = \bar{\alpha}\,\bar{A}$, for $\alpha \in C$ and $A \in M_{m\times n}(C)$

c) $\bar{\bar{A}} = A$, for $A \in M_{m\times n}(C)$

Let us see an example of obtaining the conjugate of a matrix.

Example 7: Find the conjugate of $\begin{bmatrix} 1 & i \\ 2+i & -3-2i \end{bmatrix}$

Solution: By definition, the required matrix will be

$\begin{bmatrix} 1 & -i \\ 2-i & -3+2i \end{bmatrix}$

Example 8: What is the conjugate of $\begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$?

Solution: Note that this matrix has only real entries. Thus, the complex conjugate of each entry is itself. This means that the conjugate of this matrix is itself.

This example leads us to make the following observation.

Remark: $\bar{A} = A$ if and only if A is a real matrix.

Try the following exercise now.

E 21) Calculate the conjugate of $\begin{bmatrix} i & 2 \\ 3 & i \end{bmatrix}$.

We combine what we have learnt in the previous two sub-sections now.

7.3.3 Conjugate Transpose

Given a matrix $A \in M_{m\times n}(F)$ we form a matrix B by taking the conjugate of A^t. Then $B = \bar{A}^t$, is called the **conjugate transpose of A**.

Example 9: Find $\bar{A}^t$ where $A = \begin{bmatrix} 1 & i \\ 2+i & -3-2i \end{bmatrix}$

Solution: Firstly,

$$A^t = \begin{bmatrix} 1 & 2+i \\ i & -3-2i \end{bmatrix}. \text{ Then}$$

$$\bar{A}^t = \begin{bmatrix} 1 & 2-i \\ -i & -3+2i \end{bmatrix}.$$

Now, note a peculiar occurrence. If we first calculate $\bar{A}$ and then take its transpose, we get the same matrix, namely, $\bar{A}^t$. That is, $(\bar{A})^t = \bar{A}^t$.

In general, $(\bar{A})^t = \bar{A}^t \ \forall A \in M_{m\times n}(C)$.

E 22) Show that $A = \bar{A}^t \Rightarrow A$ is a square matrix.

$\bar{A}^t$ is also denoted by A^* or A^*.

E 22 leads us to the following definitions.

Definitions: A square matrix A for which $\bar{A}^t = A$ is called a **Hermitian matrix.** A square matrix A is called a **skew-Hermitian matrix** if $\bar{A}^t = -A$.

For example, the matrix $\begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$ is Hermitian, whereas the matrix $\begin{bmatrix} i & 1+i \\ -1+i & 0 \end{bmatrix}$ is a skew-Hermitian matrix.

Note: If $A = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$, then $A = A^t = \bar{A}^t$ (since the entries are all real). $\therefore$ A is symmetric as well as Hermitian. In fact, for a real matrix A, A is Hermitian iff A is symmetric. Similarly, A is skew-Hermitian iff A is skew-symmetric.

We will now discuss two important, and often-used, types of square matrices.

7.4 SOME TYPES OF MATRICES

In this section we will define a diagonal matrix and a triangular matrix.

7.4.1 Diagonal Matrix

Let U and V be vector spaces over F of dimension n. Let $B_1 = \{e_1, \ldots, e_n\}$ and $B_2 = \{f_1, \ldots, f_n\}$ be based of U and V, respectively. Let $d_1, \ldots, d_n \in F$. Consider the transformation
$T : U \to V : T(a_1e_1 + \ldots + a_ne_n) = a_1d_1f_1 + \ldots + a_nd_nf_n$.
Then $T(e_1) = d_1f_1, T(e_2) = d_2f_2, \ldots, T(e_n) = d_nf_n$.

$$\therefore [T]_{B_1, B_2} = \begin{bmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ . & . & \ldots & . \\ 0 & 0 & \ldots & d_n \end{bmatrix}$$

Such a matrix is called a diagonal matrix. Let us see what this means.

Let $A = [a_{ij}]$ be a square matrix. The entries $a_{11}, a_{22}, \ldots, a_{nn}$ are called the **diagonal entries** of A. This is because they lie along the diagonal, from left to right, of the matrix. All the other entries of A are called the off-diagonal entries of A.

A square matrix whose off-diagonal entries are zero (i.e., $a_{ij} = 0 \ \forall i \neq j$) is called a **diagonal matrix.** The diagonal matrix

$$\begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix}$$

is denoted by diag $(d_1, d_2, \ldots\ldots, d_n)$.

Note: The d_i's may or may not be zero. What happens if all the d_i's are zero? Well, we get the $n \times n$ zero matrix, which corresponds to the zero operator.

If $d_i = 1 \ \forall i = 1, \ldots\ldots, n$, we get the **identity matrix**, I_n (or I, when the size is understood).

E 23) Show that I_n is the matrix associated to the identity operator from R^n to R^n.

If $\alpha \in F$, the linear operator $\alpha I: R^n \to R^n : \alpha I(v) = \alpha v$, for all $v \in R^n$, is called a **scalar operator.** Its matrix with respect to any basis is $\alpha I = \text{diag}(\alpha, \alpha, \ldots\ldots, \alpha)$. Such a matrix is called a **scalar matrix.** It is a diagonal matrix whose diagonal entries are all equal.

With this much discussion on diagonal matrices, we move onto describe triangular matrices.

7.4.2 Triangular Matrix

Let $B = \{e_1, e_2, \ldots, e_n\}$ be a basis of a vector space V. Let $S \in L(V,V)$ be an operator such that

$S(e_1) = a_{11}e_1$

$S(e_2) = a_{12}e_1 + a_{22}e_2$

$\vdots \quad \vdots \quad \vdots$

$S(e_n) = a_{1n}\, e_1 + a_{2n}\, e_2 + \ldots + a_{nn}\, e_n$,

Then, the matrix of S with respect to B is

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ 0 & a_{22} & \ldots & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \ldots & \vdots \\ 0 & 0 & 0 & \ldots & a_{nn} \end{bmatrix}$$

Note that $a_{ij} = 0 \; \forall \; i > j$

A square matrix A such that $a_{ij} = 0 \; \forall \; i > j$ is called an **upper triangular matrix.** If $a_{ij} = 0 \; \forall \; i \geq j$, then A is called **strictly upper triangular.**

For example, $\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are all upper triangular, while $\begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix}$ is strictly upper triangular.

Note that every strictly upper triangular matrix is an upper triangular matrix.

Now let $T : V \to V$ be an operator such that $T(e_j)$ is a linear combination of $e_j, e_{j+1}, \ldots, e_n \; \forall j$.
The matrix of T with respect to B is

$$[T]_B = \begin{bmatrix} b_{11} & 0 & 0 & \ldots & 0 \\ b_{21} & b_{22} & \vdots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \ldots & b_{nn} \end{bmatrix}$$

Note that $b_{ij} = 0 \; \forall \; i < j$

Such a matrix is called a **lower triangular matrix.** If $b_{ij} = 0$ for all $i \leq j$, **then B is said to be a strictly lower triangular matrix.**

The matrix

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 1 & 0 & 5 & 0 \end{bmatrix}$$

is a strictly lower triangular matrix. Of course, it is also lower triangular!

Remark: If A is an upper triangular 3 x 3 matrix, say

$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$, then $A^t = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{bmatrix}$, a lower triangular matrix.

In fact, for any $n \times n$ upper triangular matrix A, its transpose is lower triangular, and vice versa.

E 24) If an upper triangular matrix A is symmetric, then show that it must be a diagonal matrix.

E 25) Show that the diagonal entries of a skew-symmetric matrix are all-zero, but the converse is not true.

Let us now see how to define the product of two or more matrices.

7.5 MATRIX MULTIPLICATION

We have already discussed scalar multiplication. Now we see how to multiply two matrices. Again, the motivation for this operation comes from linear transformations.

7.5.1 Matrix of the Composition of Linear Transformations

Let U, V and W be vector spaces over F, of dimensions p, n and m, respectively. Let B_1, B_2 and B_3 be bases of these respective spaces. Let $T \in L(U,V)$ and $S \in L(V,W)$. Then $ST (= S \circ T) \in L(U, W)$ (see Sec. 6.4).

Suppose $[T]_{B_1, B_2} = B = [b_{jk}]_{n \times p}$

and $[S]_{B_2, B_3} = A = [a_{ij}]_{m \times n}$

We ask : What is the matrix $[ST]_{B_1, B_3}$?

To answer this we suppose

$B_1 = \{e_1, e_2, \ldots\ldots, e_p\}$

$B_2 = \{f_1, f_2, \ldots\ldots, f_n\}$

$B_3 = \{g_1, g_2, \ldots\ldots, g_m\}$

Then, we know that $T(e_k) = \sum_{j=1}^{n} b_{jk} f_j \ \forall\ k = 1, 2, \ldots, p.$

and $S(f_j) = \sum_{i=1}^{m} a_{ij} g_i \ \forall\ j = 1, 2, \ldots, n.$

Therefore, $S \circ T (e_k) = S(T(e_k)) = S\left(\sum_{j=1}^{n} b_{jk} f_j\right) = b_{1k} S(f_1) + b_{2k} S(f_2) + \cdots + b_{nk} S(f_n)$

$$= b_{1k}\left(\sum_{i=1}^{m} a_{i1} g_i\right) + b_{2k}\left(\sum_{i=1}^{m} a_{i2} g_i\right) + \cdots + b_{nk}\left(\sum_{i=1}^{m} a_{in} g_i\right)$$

$$= \sum_{i=1}^{m} (a_{i1} b_{1k} + a_{i2} b_{2k} + \cdots + a_{in} b_{nk}) g_i,$$ on collecting the coefficients of g_i.

Thus, $[ST]_{B_1, B_3} = [c_{ik}]_{m \times p}$, where $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$.

We define the matrix $[c_{ik}]$ to be the product AB.

So, let us see how we obtain AB from A and B.

Let $A = [a_{ij}]_{m \times n}$, $B = [b_{jk}]_{n \times p}$ be two matrices over F of sizes $m \times n$ and $n \times p$, **respectively.** We define AB to be the $m \times p$ matrix C whose (i,k)th entry is

The product of an $m \times n$ and an $n \times p$ matrix is an $m \times p$ matrix.

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1} b_{1k} + a_{i2} b_{2k} + \cdots + a_{in} b_{nk}$$

In order to obtain the (i,k) th element of AB, take the ith row of A and the kth column of B

They are both n-tuples. Multiply their corresponding elements and add up all these products.

For example, if the 2nd row of A = [1 2 3], and the 3rd column of

$B=\begin{bmatrix}4\\5\\6\end{bmatrix}$, then the (2, 3) entry of AB = $1\times 4+2\times 5+3\times 6=32$.

Note that two matrices A and B can only be multiplied if the number of columns of A = the number of rows of B. The following illustration may help in explaining what we do to obtain the product of two matrices.

$$\overset{\textbf{A}}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & & \vdots\\ a_{i1} & a_{i2} & \cdots & a_{in}\\ \vdots & \vdots & & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}} \overset{\textbf{B}}{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} & \cdots & b_{1p}\\ b_{21} & b_{22} & \cdots & b_{2k} & \cdots & b_{2p}\\ \vdots & \vdots & & \vdots & & \vdots\\ \vdots & \vdots & & \vdots & & \vdots\\ \vdots & \vdots & & \vdots & & \vdots\\ b_{n1} & b_{n2} & \cdots & b_{nk} & \cdots & b_{np}\end{bmatrix}} = \overset{\textbf{AB}}{\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1k} & \cdots & c_{1p}\\ c_{21} & c_{22} & \cdots & c_{2k} & \cdots & c_{2p}\\ \vdots & \vdots & & \vdots & & \vdots\\ c_{i1} & c_{i2} & \cdots & c_{ik} & \cdots & c_{ip}\\ \vdots & \vdots & & \vdots & & \vdots\\ c_{m1} & c_{m2} & \cdots & c_{mk} & \cdots & c_{mp}\end{bmatrix}}$$

where $c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$

Note: This is a very new kind of operation so take your time in trying to understand it.

To get you used to matrix multiplication we consider the product of a row and a column matrix.

Let $A = [a_1, a_2, \ldots, a_n]$ be a $1\times n$ matrix and $B=\begin{bmatrix}b_1\\b_2\\\vdots\\b_n\end{bmatrix}$ be an $n\times 1$ matrix. Then AB is the 1×1 matrix

$[a_1 b_1 + a_2 b_2 + \cdots + a_n b_n]$.

E 26) What is $[1 \quad 0 \quad 0]\begin{bmatrix}1\\2\\3\end{bmatrix}$?

Now for another example.

Example 10: Let $A = \begin{bmatrix}1 & 0 & 0\\7 & 0 & 8\\0 & 0 & 9\end{bmatrix}$, $B = \begin{bmatrix}2 & 1\\3 & 5\\4 & 0\end{bmatrix}$

Find AB, if it is defined.

Solution: AB is defined because the number of columns of A = 3 = number of rows of B.

$$AB = \begin{bmatrix}1.2 + 0.3 + 0.4 & 1.1 + 0.5 + 0.0\\ 7.2 + 0.3 + 8.4 & 7.1 + 0.5 + 8.0\\ 0.2 + 0.3 + 9.4 & 0.1 + 0.5 + 9.0\end{bmatrix} = \begin{bmatrix}2 & 1\\46 & 7\\36 & 0\end{bmatrix}$$

Notice that BA is not defined because the number of columns of B = 2 ≠ number of rows of A. Thus, **if AB is defined then BA may not be defined.**

In fact, even if AB and BA are both defined it is possible that AB ≠ BA. Consider the following example.

Example 11: Let $A = \begin{bmatrix}1 & 1 & 0\\0 & 1 & 1\end{bmatrix}$, $B = \begin{bmatrix}0 & 1\\1 & 1\\1 & 1\end{bmatrix}$. Is AB = BA?

Solution: AB is a 2×2 matrix. BA is a 3×3 matrix.

So AB and BA are both defined, but they are of different sizes. Thus, $AB \neq BA$.

Another point of difference between multiplication of numbers and matrix multiplication is that $A \neq 0$, $B \neq 0$, but AB can be zero.

For example, if $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$,

then $AB = \begin{bmatrix} 1 \times 1 + 1(-1) & 1 \times 0 + 1 \times 0 \\ 1 \times 1 + 1(-1) & 1 \times 0 + 1 \times 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$.

So you see, the product of two non-zero matrices can be zero.

The following exercises will give you some practice in matrix multiplication.

E E 27) Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$.

Write AB and BA, if defined.

E E 28) Let $C = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $D = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}$.

Write C + D, CD and DC, if defined. Is CD = DC?

E E 29) With A, B as in E 27, calculate $(A + B)^2$ and $A^2 + 2AB + B^2$. Are they equal? (Here A^2 means A.A.)

E E 30) Let $A = \begin{bmatrix} -bd & b \\ -d^2 b & db \end{bmatrix}$, $b, d \in F$. Find A^2.

E 31) Calculate $\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}$ and $[x\ y\ z]\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 3\end{bmatrix}$

E 32) Take a 3 × 2 matrix A whose end row consists of zeros only. Multiply it by any 2 × 4 matrix B. Show that the 2nd row of AB consists of zeros only. (In fact, for any two matrices A and B such that AB is defined, if the ith row of A is the zero vector, then the ith row of AB is also the zero vector. Similarly, if the jth column of B is the zero vector, then the jth column of AB is the zero vector.)

We now make an observation.

Remark: If $T \in L(U, V)$ and $S \in L(V, W)$, then

$[ST]_{B_1, B_3} = [S]_{B_2, B_3}\ [T]_{B_1, B_2}$, where B_1, B_2, B_3 are the bases of U, V, W, respectively.

Let us illustrate this remark.

Example 12: Let $T : R^2 \rightarrow R^3$ be a linear transformation such that T (x, y) = (2x + y, x + 2y, x + y). Let $S : R^3 \rightarrow R^2$ be defined by S (x, y, z) = (−y + 2z, y−z). Obtain the matrices $[T]_{B_1, B_2}$, $[S]_{B_2, B_1}$, and $[SoT]_{B_1}$, and verify that $[SoT]_{B_1} = [S]_{B_2, B_1}\ [T]_{B_1, B_2}$, where B_1 and B_2 are the standard bases in R^2 and R^3, respectively.

Solution : Let $B_1 = \{e_1, e_2\}$, $B_2 = \{f_1, f_2, f_3\}$.

Then $T(e_1) = T(1,0) = (2,1,1) = 2f_1 + f_2 + f_3$

$T(e_2) = T(0,1) = (1,2,1) = f_1 + 2f_2 + f_3$

Thus,

$$[T]_{B_1, B_2} = \begin{bmatrix}2 & 1\\1 & 2\\1 & 1\end{bmatrix}$$

Also,

$S(f_1) = S(1,0,0) = (0,0) = 0.e_1 + 0.e_2$

$S(f_2) = S(0,1,0) = (-1,1) = -e_1 + e_2$

$S(f_3) = S(0,0,1) = (2,-1) = 2e_1 - e_2$

Thus,

$$[S]_{B_2, B_1} = \begin{bmatrix}0 & -1 & 2\\0 & 1 & -1\end{bmatrix}.$$

$$\text{So, } [S]_{B_2, B_1}[T]_{B_1, B_2} = \begin{bmatrix}0 & -1 & 2\\0 & 1 & -1\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 2\\1 & 1\end{bmatrix}$$

$$= \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} = I_2$$

Also, SoT (x,y) = S (2x + y, x + 2y, x + y)

= (− x − 2y + 2x + 2y, x + 2y − x − y)

= (x,y)

Thus, $S \circ T = I$, the identity map.

This means $[S \circ T]_{B_1} = I_3$

Hence, $[S \circ T]_{B_1} = [S]_{B_2,B_1} [T]_{B_1,B_2}$

E 33) Let $S : \mathbf{R}^3 \to \mathbf{R}^3 : S(x, y, z) = (0, x, y)$, and $T : \mathbf{R}^3 \to \mathbf{R}^3 : T(x, y, z) = (x, 0, y)$
Show that $[S \circ T]_B = [S]_B [T]_B$, where B is the standard basis of $\mathbf{R}^3$.

We will now look a little closer at matrix multiplication.

7.5.2 Properties of a Matrix Product

We will now state 5 properties concerning matrix multiplication. (Their proofs could get a little technical, and we prefer not to give them here.)

(1) **Associative Law:** If, A, B, C are $m \times n$, $n \times p$ and $p \times q$ matrices, respectively, over F, then (AB) C = A (BC), i.e., matrix multiplication is associative.

(2) **Distributive Law:** If A is an $m \times n$ matrix and B, C are $n \times p$ matrices, then A (B + C) = AB + AC.
Similarly if, A and B are $m \times n$ matrices, and C is an $n \times p$ matrix, then
(A + B)C = AC + BC.

(3) **Multiplicative Identity:** In Sec. 7.4.1, we defined the identity matrix I_n. This acts as the multiplicative identity for matrix multiplication. We have $AI_n = A$, $I_m A = A$, for every $m \times n$ matrix A.

(4) If $\alpha \in \mathbf{F}$, and A, B are $m \times n$ and $n \times p$ matrices over **F**, respectively, then $\alpha(\mathbf{AB}) = (\alpha A) B = A (\alpha B)$.

(5) If A, B are $m \times n$, $n \times p$ matrices over F, respectively, then $(\mathbf{AB})^t = \mathbf{B}^t \mathbf{A}^t$. (This says that the operation of taking the transpose of a matrix is anti-commutative.)
These properties can help you in solving the following exercises.

E 34) Show that $(A + B)^2 = A^2 + AB + BA + B^2$, for any two $n \times n$ matrices A and B.

E 35) For $A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ -3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -2 & -5 \\ 3 & 4 & 0 \end{bmatrix}$, show that 2 (AB) = (2A) B.

E 36) Let $A = \begin{bmatrix} 2 & -1 & 0 \\ 1 & 0 & -3 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -4 & 0 \\ 2 & -1 & 3 \\ 4 & 0 & -2 \end{bmatrix}$

Find $(AB)^t$ and $B^t A^t$. Are they equal?

E 37) Let A, B be two symmetric $n \times n$ matrices over F. Show that AB is symmetric if and only if AB = BA.

The following exercise is a nice property of the product of diagonal matrices.

E 38) Let A, B be two diagonal $n \times n$ matrices over F. Show that AB is also a diagonal matrix.

Now we shall go on to introduce you to the concept of an invertible matrix.

7.6 INVERTIBLE MATRICES

In this section we will first explain what invertible matrices are. Then we will see what we mean by the matrix of a change of basis. Finally, we will show you that such a matrix must be invertible.

7.6.1 Inverse of a Matrix

Just as we defined the operations on matrices by considering them on linear operators first, we give a definition of invertibility for matrices based on considerations of invertibility of linear operators.

It may help you to recall what we mean by an invertible linear transformation. A linear transformation $T : U \to V$ is invertible if

a) T is 1 – 1 and onto, or, equivalently,

b) there exists a linear transformation $S : V \to U$ such that $S \circ T = I_U$, $T \circ S = I_V$.

In particular, $T \in L(V, V)$ is said to be invertible if $\exists\, S \in L(V, V)$ such that ST = TS = I.

We have the following theorem involving the matrix of an invertible linear operator.

Theorem 4: Let V be an n-dimensional vector space over a field F, and B be a basis of V. Let $T \in L(V, V)$. T is invertible iff there exists $A \in M_n(F)$ such that $[T]_B A = I_n = A[T]_B$

Proof : Suppose T is invertible Then $\exists S \in L(V, V)$ such that $TS = ST = I$. Then, by Theorem 2, $[TS]_B = [ST]_B = I$. That is, $[T]_B [S]_B = [S]_B [T]_B = I$. Take $A = [S]_B$. Then $[T]_B A = I = A[T]_B$.

Conversely, suppose $\exists$ a matrix A such that $[T]_B A = A[T]_B = I$.

Let $S \in L(V, V)$ be such that $[S]_B = A$. (S exists because of Theorem 2.) Then $[T]_B [S]_B = [S]_B [T]_B = I = [I]_B$. Thus, $[TS]_B = [ST]_B = [I]_B$.

So, by Theorem 2, $TS = ST = I$. That is, T is invertible.

Theorem 4 motivates us to give the following definition.

Definition : A matrix $A \in M_n(F)$ is said to be invertible if $\exists B \in M_n(F)$ such that $AB = BA = I_n$.

Remember, only a square matrix can be invertible.

I_n is an example of an invertible matrix, since $I_n \cdot I_n = I_n$. On the other hand, the $n \times n$ zero matrix 0 is not invertible, since $0A = 0 \neq I_n$, for any A.

Note that Theorem 4 says that **T is invertible iff $[T]_B$ is invertible.**

We give another example of an invertible matrix now.

Example 13: Is $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ invertible?

Solution: Suppose A were invertible. Then $\exists B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that $AB = I = BA$. Now,

$$AB = I \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} a+c & b+d \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow c = 0,\ d = 1,\ a = 1,\ b = -1$$

$$\therefore B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$ Now you can also check that $BA = I$.

Therefore A is invertible.

We now show that if an inverse of a matrix exists, it must be unique.

Theorem 5 : Suppose $A \in M_n(F)$ is invertible. There exists a unique matrix $B \in M_n(F)$ such that $AB = BA = I$.

Proof: Suppose $B, C \in M_n(F)$ are two matrices such that $AB = BA = I$, and $AC = CA = I$. Then $B = BI = B(AC) = (BA)C = IC = C$.

Because of Theorem 5 we can make the following definition.

Definition : Let A be an invertible matrix. The unique matrix B such that $AB = BA = I$ is called the inverse of A and is denoted by A^{-1}.

Let us take an example.

Example 14: Calculate the product AB, where

$$A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix}.$$

Use this to calculate A^{-1}.

Solution : Now $AB = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & a+b \\ 0 & 1 \end{bmatrix}$.

Now, how can we use this to obtain A^{-1}? Well, if $AB = I$, then $a+b = 0$. So, if we take

$$B = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix},$$

we get $AB = BA = I$. Thus, $A^{-1} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix}$

E 39) Is the matrix $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$ invertible? If so, find its inverse.

We will now make a few observations about the matrix inverse, in the form of a theorem.

Theorem 6: a) If A is invertible, then

i) A^{-1} is invertible and $(A^{-1})^{-1} = A$.

ii) A^t is invertible and $(A^t)^{-1} = (A^{-1})^t$.

b) If $A, B \in M_n(F)$ are invertible, then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$

Proof: (a) By definition,

$A A^{-1} = A^{-1} A = I$ ....(1)

i) Equation (1) shows that A^{-1} is invertible and $(A^{-1})^{-1} = A$.

ii) If we take transposes in Equation (1) and use the property that $(AB)^t = B^t A^t$, we get $(A^{-1})^t A^t = A^t (A^{-1})^t = I^t = I$.

So A^t is invertible and $(A^t)^{-1} = (A^{-1})^t$.

(b) To prove this we will use the associativity of matrix multiplication. Now
$(AB)(B^{-1}A^{-1}) = [A(BB^{-1})]A^{-1} = A A^{-1} = I$.
$(B^{-1}A^{-1})(AB) = B^{-1}[(A^{-1}A)B] = B^{-1}B = I$.

So AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$.

We now relate matrix invertibility with the linear independence of its rows or columns. When we say that the m rows of $A = [a_{ij}] \in M_{m\times n}(F)$ are linearly independent, what do we mean? Let $R_1, \ldots, R_m$ be the m row vectors $[a_{11}, a_{12}, \ldots, a_{1n}], [a_{21}, \ldots, a_{2n}], \ldots, [a_{m1}, \ldots, a_{mn}]$, respectively. We say that they are linearly independent if, whenever $\exists\, a_1, \ldots, a_m \in F$ such that $a_1R_1 + \ldots + a_m R_m = 0$,

then $a_1 = 0, \ldots\ldots, a_m = 0$.

Similarly, the n columns $C_1, \ldots, C_n$ of A are linearly independent if $b_1C_1 + \ldots b_n C_n = 0$
$\Rightarrow b_1 = 0, b_2 = 0, \ldots, b_n = 0$, where $b_1, \ldots, b_n \in F$.

We have the following result.

Theorem 7: Let $A \in M_n(F)$. Then the following conditions are equivalent.

a) A is invertible.

b) The columns of A are linearly independent.

c) The rows of A are linearly independent.

Proof: We first prove (a) $\Leftrightarrow$ (b), using Theorem 4. Let V be an n-dimensional vector space over F and $B = \{e_1, \ldots, e_n\}$ be a basis of V. Let $T \in L(V,V)$ be such that $[T]_B = A$. Then A is invertible iff T is invertible iff $T(e_1), T(e_2), \ldots, T(e_n)$ are linearly independent (see Unit 5, Theorem 9). Now we define the map

$$\theta: V \to M_{n\times 1}(F): \theta(a_1 e_1 + \cdots + a_n e_n) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Before continuing the proof we give an exercise.

E 40) Show that θ is a well-defined isomorphism.

Now let us go on with proving Theorem 7.

Let $C_1, C_2, \ldots, C_n$ be the columns of A. Then $\theta(T(e_i)) = C_i$ for all $i = 1, \ldots, n$. Since θ is an isomorphism, $T(e_1), \ldots, T(e_n)$ are linearly independent iff $C_1, C_2, \ldots, C_n$ are linearly independent. Thus, A is invertible iff $C_1, \ldots, C_n$ are linearly independent. Thus, we have proved (a) $\Leftrightarrow$ (b).

Now, the equivalence of (a) and (c) follows because A is invertible $\Leftrightarrow$ A^t is invertible

$\Leftrightarrow$ the columns of A^t are linearly independent (as we have just shown)

$\Leftrightarrow$ the rows of A are linearly independent (since the columns of A^t are the rows of A).

So we have shown that (a) $\Leftrightarrow$ (c).

Thus, the theorem is proved.

From the following example you can see how Theorem 7 can be useful.

Example 15:

Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \in M_3(\mathbf{R})$.

Determine whether or not A is invertible.

Solution: Let R_1, R_2, R_3 be the rows of A. We will show that they are linearly independent.

Suppose $xR_1 + yR_2 + zR_3 = 0$, where $x, y, z \in \mathbf{R}$. Then,

$x(1,0,1) + y(0,1,1) + z(1,1,1) = (0, 0, 0)$. This gives us the following equations.

$x + z = 0$

$y + z = 0$

$x + y + z = 0$

On solving these we get $x = 0, y = 0, z = 0$.

Thus, by Theorem 7, A is invertible.

E 41) Check if

$\begin{bmatrix} 2 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 3 & 0 \end{bmatrix} \in M_3(\mathbf{Q})$ is invertible.

We will now see how we associate a matrix to a change of basis. This association will be made use of very often in the next block.

7.6.2 Matrix of Change of Basis

Let V be an n-dimensional vector space over F. Let $B = \{e_1, e_2, \ldots, e_n\}$ and $B' = \{e_1', e_2', \ldots, e_n'\}$ be two bases of V. Since $e_j' \in V$, for every j, it is a linear combination of the elements of B. Suppose,

$$e_j' = \sum_{i=1}^{n} a_{ij} e_i \ \forall j = 1, \ldots, n.$$

The n×n matrix $A = [a_{ij}]$ is called the **matrix of the change of basis from B to B'**. It is denoted by $M_B^{B'}$.

Note that A is the matrix of the transformation $T \in L(V, V)$ such that $T(e_j) = e_j' \ \forall j = 1, \ldots, n$, with respect to the basis B. Since $\{e_1', \ldots, e_n'\}$ is a basis of V, from Unit 5 we see that T is 1–1 and onto. Thus T is invertible. So A is invertible. Thus, **the matrix of the change of basis from B to B' is invertible.**

Note: a) $M_B^B = I_n$. This is because, in this case $e_i' = e_i \ \forall i = 1, 2, \ldots, n$.

b) $M_B^{B'} = [I]_{B', B}$. This is because

$$I(e_j') = e_j' = \sum_{i=1} a_{ij} e_i \ \forall j = 1, 2, \ldots, n$$

Now suppose A is any invertible matrix. By Theorem 2, $\exists\, T \in L(V, V)$ such that $[T]_B = A$. Since A is invertible, T is invertible. Thus, T is 1–1 and onto. Let $f_i = T(e_i) \ \forall i = 1, 2, \ldots, n$. Then $B' = \{f_1, f_2, \ldots, f_n\}$ is also a basis of V, and the matrix of change of basis from B to B' is A.

In the above discussion, we have just proved the following theorem.

Theorem 8: Let $B = \{e_1, e_2, \ldots, e_n\}$ be a fixed basis of V. The mapping $B' \to M_B^{B'}$ is a 1–1 and onto correspondence between the set of all bases of V and the set of invertible n×n matrices over F.

Let us see an example of how to obtain $M_B^{B'}$

Example 16: In R^2, $B = \{e_1, e_2\}$ is the standard basis. Let B' be the basis obtained by rotating B through an angle θ in the anti-clockwise direction (see Fig. 1). Then $B' = (e_1', e_2')$ where $e_1' = (\cos\theta, \sin\theta)$, $e_2' = (-\sin\theta, \cos\theta)$. Find $M_B^{B'}$

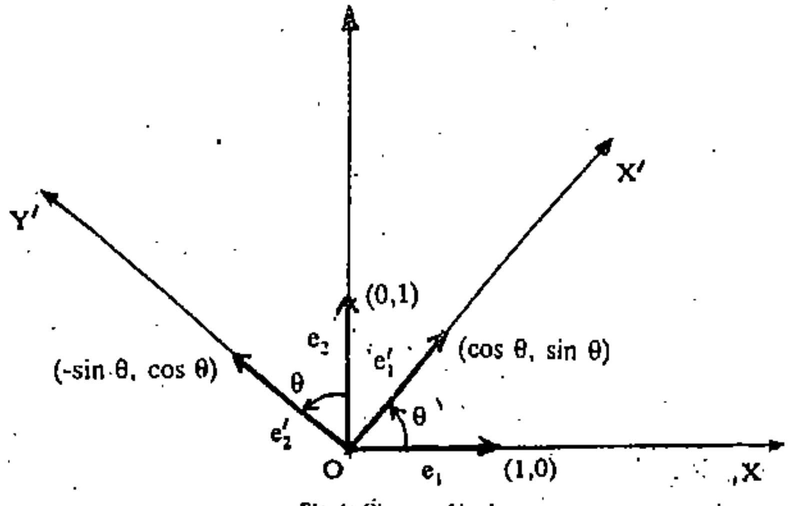


Fig. 1: Change of basis.

Solution : $e_1' = \cos\theta\,(1,0) + \sin\theta\,(0,1)$, and

$e_2' = -\sin\theta(1, 0) + \cos\theta\,(0,1)$

Thus, $M_B^{B'} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Try the following exercise.

E 42) Let B be the standard basis of R^3 and B' be another basis such that $M_B^{B'} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$
What are the elements of B'?

What happens if we change the basis more than once? The following theorem tells us something about the corresponding matrices.

Theorem 9: Let B, B′, B″ be three bases of V. Then $M_{B}^{B'} M_{B'}^{B''} = M_{B}^{B''}$

Proof: Now, $M_{B}^{B'} M_{B'}^{B''} = [I]_{B',B}[I]_{B'',B'}$

$= [I \circ I]_{B'',B} = M_{B}^{B''}$

An immediate useful consequence is

Corollary : Let B, B′ be two bases of V. Then $M_{B}^{B'} M_{B'}^{B} = I = M_{B'}^{B} M_{B}^{B'}$

That is, $\left(M_{B}^{B'}\right)^{-1} = M_{B'}^{B}$

Proof : By Theorem 9,

$M_{B}^{B'} M_{B'}^{B} = M_{B}^{B} = I$

Similarly, $M_{B'}^{B} M_{B}^{B'} = M_{B'}^{B'} = I$.

But, how does the change of basis affect the matrix associated to a given linear transformation? In Sec. 7.2 we remarked that the matrix of a linear transformation depends upon the pair of bases chosen. The relation between the matrices of a transformation with respect to two pairs of bases can be described as follows.

Theorem 10: Let $T \in L(U, V)$. Let $B_1 = \{e_1, \ldots, e_n\}$ and $B_2 = \{f_1, \ldots, f_m\}$ be a pair of bases of U and V, respectively.

Let $B_1' = \{e_1', \ldots, e_n'\}$, $B_2' = \{f_1', \ldots, f_m'\}$ be another pair of bases of U and V, respectively. Then.

$[T]_{B_1', B_2'} = M_{B_2'}^{B_2} [T]_{B_1, B_2} M_{B_1}^{B_1'}$

Proof: $[T]_{B_1', B_2'} = [I_V \circ T \circ I_U]_{B_1', B_2'} = [I_V]_{B_2, B_2'} [T]_{B_1, B_2} [I_U]_{B_1', B_1}$

(where I_U = identity map on U and I_V = identity map on V)

$= M_{B_2'}^{B_2} [T]_{B_1, B_2} M_{B_1}^{B_1'}$

Now, a corollary to Theorem 10, which will come in handy in the next block.

Corollary: Let $T \in L(V, V)$ and B, B′ be two bases of V. Then $[T]_{B'} = P^{-1} [T]_B P$, where $P = M_{B}^{B'}$

Proof: $[T]_{B'} = M_{B'}^{B} [T]_B M_{B}^{B'} = P^{-1} [T]_B P$, by the corollary to Theorem 9.

Let us now recapitulate all that we have covered in this unit.

7.7 SUMMARY

We briefly sum up what has been done in this unit.

1) We defined matrices and explained the method of associating matrices with linear transformations.
2) We showed what we mean by sums of matrices and multiplication of matrices by scalars.
3) We proved that $M_{m \times n}(F)$ is a vector space of dimension mn over F.
4) We defined the transpose of a matrix, the conjugate of a complex matrix, the conjugate transpose of a complex matrix, a diagonal matrix, identity matrix, scalar matrix and lower and upper triangular matrices.

5) We defined the multiplication of matrices and showed its connection with the composition of linear transformations. Some properties of the matrix product were also listed and used.

6) The concept of an invertible matrix was explained.

7) We defined the matrix of a change of basis, and discussed the effect of change of bases on the matrix of a linear transformation.

7.8 SOLUTIONS/ANSWERS

E1) a) You want the elements in the 1st row and the 2nd column. They are 2 and 5, respectively

b) [0 0 7]

c) The second column of A is $\begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}$

The first column of B is also $\begin{bmatrix} 2 \\ 5 \\ 0 \end{bmatrix}$

d) B only has 3 rows. Therefore, there is no 4th row of B.

E2) They are infinitely many answers. We give

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix} \text{and} \begin{bmatrix} 1 & 0 \\ 2 & 0 \\ 3 & 0 \\ 4 & 0 \end{bmatrix}$$

E3) No. Because they are of different sizes.

E4) Suppose $B_1' = \{(1, 0, 1), (0, 2, -1), (1, 0, 0)\}$ and $B_2' = \{(0, 1), (1, 0)\}$

Then $T(1, 0, 1) = (1,0) = 0.(0,1) + 1.(1, 0)$

$T(0, 2, -1) = (0,2) = 2.(0,1) + 0.(1, 0)$

$T(1, 0, 0) = (1,0) = 0.(0,1) + 1.(1,0)$.

$$\therefore [T]_{B_1', B_2'} = \begin{bmatrix} 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

E5) $B_1 = \{e_1, e_2, e_3\}$, $B_2 = \{f_1, f_2\}$ are the standard bases (given in Example 3).

$T(e_1) = T(1, 0, 0) = (1, 2) = f_1 + 2f_2$

$T(e_2) = T(0, 1, 0) = (2, 3) = 2f_1 + 3f_2$

$T(e_3) = T(0, 0, 1) = (2, 4) = 2f_1 + 4f_2$

$$\therefore [T]_{B_1, B_2} = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 3 & 4 \end{bmatrix}$$

E6) $T(1,0,0) = (1,2) = 1.(1,2) + 0.(2,3)$

$T(0, 1, 0) = (2, 3) = 0.(1, 2) + 1.(2, 3)$

$T(1, -2, 1) = (-1, 0) = 3(1,2) - 2(2, 3)$

$$\therefore [T]_{B_1', B_2'} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -2 \end{bmatrix}$$

E7) Let $B = \{1, t, t^2, t^3\}$. Then

$D(1) = 0 = 0.1 + 0.t + 0.t^2 + 0.t^3$

$D(t) = 1 = 1.1 + 0.t + 0.t^2 + 0.t^3$

$D(t^2) = 2t = 0.1 + 2.t + 0.t^2 + 0.t^3$

$D(t^3) = 3t^2 = 0.1 + 0.t + 3.t^2 + 0.t^3$.

Therefore $[D]_B$ is the given matrix.

E8) We know that

$T(e_1) = f_1$

$T(e_2) = f_1 + f_2$

$T(e_3) = f_2$

Therefore, for any $(x, y, z) \in \mathbb{R}^3$,

$T(x,y,z) = T(xe_1 + ye_2 + ze_3) = xT(e_1) + yT(e_2) + zT(e_3)$.

$= xf_1 + y(f_1 + f_2) + zf_2 = (x + y) f_1 + (y + z)f_2$

$= (x + y, y + z)$

That is, $T:\mathbb{R}^3 \to \mathbb{R}^2: T(x,y,z) = (x + y, y + z)$.

E9) We are given that

$T(1) = 0.1 + 1.i = i$

$T(i) = (-1).1 + 0.i = -1$

$\therefore$, for any $a+ib \in \mathbb{C}$, we have

$T(a+ib) = aT(1) + bT(i) = ai - b$

E 10) a) Since $[1\ 2]$ is of size 1×2 and $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ is of size 2×1,

the sum of these matrices is not defined.

b) Both matrices are of the same size, namely, 2×2. Their sum is the matrix

$$\begin{bmatrix} 1 + (-1) & 0 + 0 \\ 0 + 0 & 1 + (-1) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

E11) $3\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$, $3\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

and $3\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 3\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \end{bmatrix}$

Notice that $3\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix}$

E12) $B_1 = \{(1,0), (0,1)\}$, $B_2 = \{(1,0,0), (0,1,0), (0,0,1)\}$

Now $S(1,0) = (1,0,0)$

$S(0,1) = (0,0,1)$

$\therefore [S]_{B_1,B_2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$, a 3×2 matrix

Again, $T(1,0) = (0,1,0)$

$T(0,1) = (0,0,1)$

$\therefore [T]_{B_1,B_2} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$, a 3×2 matrix

$$\therefore [S + T]_{B_1,B_2} = [S]_{B_1,B_2} + [T]_{B_1,B_2} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

and $[\alpha S]_{B_1,B_2} = \alpha [S]_{B_1,B_2} = \alpha \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & 0 \\ 0 & \alpha \end{bmatrix}$, for any $\alpha \in \mathbb{R}$

E13) We will prove (v) and (vi) here. You can prove (vii) and (viii) in a similar way.

v) $\alpha (A + B) = \alpha([a_{ij}] + [b_{ij}]) = \alpha [a_{ij} + b_{ij}] = [\alpha a_{ij} + \alpha b_{ij}]$

$= [\alpha a_{ij}] + [\alpha b_{ij}] = \alpha A + \alpha B.$

vi) Prove it using the fact that $(\alpha+\beta)a_{ij} = \alpha a_{ij} + \beta a_{ij}$.

E14) Since dim $M_{2\times 3}(\mathbb{R})$ is 6, any linearly independent subset can have 6 elements, at most.

E 15) Let $\alpha, \beta \in \mathbf{R}$ such that $\alpha[1,0] + \beta[1,-1] = [0,0]$.

Then $[\alpha+\beta, -\beta] = [0,0]$. Thus, $\beta = 0$, $\alpha = 0$.

$\therefore$ the matrices are linearly independent.

E16)

$$E_{11} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \quad E_{12} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}, \text{ and so on.}$$

Now any $m \times n$ matrix $A = [a_{ij}] = a_{11}E_{11} + a_{12}E_{12} + \ldots + a_{mn}E_{mn}$ (For example, in the 2×2 situation,

$$\begin{bmatrix} 2 & 3 \\ 0 & 1 \end{bmatrix} = 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}).$$

Thus, $\{E_{ij} \mid i = 1, \ldots, m, j = 1, \ldots, n\}$ generates $M_{m\times n}(F)$. Also, if α_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$, be scalars such that $\alpha_{11}E_{11} + \alpha_{12}E_{12} + \ldots + \alpha_{mn}E_{mn} = 0$.

Then,

$$\text{we get} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

Therefore, $\alpha_{ij} = 0 \; \forall i, j$.

Hence, the given set is linearly independent. $\therefore$ it is a basis of $M_{m\times n}(F)$. The number of elements in this basis is mn.

$\therefore \dim M_{m\times n}(R) = mn$.

E17) $A^t = \begin{bmatrix} 1 & 2 \\ 2 & 0 \end{bmatrix}$. In this case $A^t = A$.

E18) b) $\alpha A = [\alpha a_{ij}]$. $\therefore (\alpha A)^t = [b_{ij}]$, where

b_{ij} = (j,i)th element of $\alpha A = \alpha a_{ji}$

= α times the (j,i)th element of A

= α times the (i,j) th element of A^t

= (i,j)th element of αA^t.

$\therefore (\alpha A)^t = \alpha A^t$.

c) Let $A = [a_{ij}]$. Then $A^t = [b_{ij}]$, where $b_{ij} = a_{ji}$.

$\therefore (A^t)^t = [c_{ij}]$, where $c_{ij} = b_{ji} = a_{ij}$

$\therefore (A^t)^t = A$.

E19) Let A be an $m \times n$ matrix. Then A^t is an $n \times m$ matrix.

$\therefore$, for $A = A^t$, their sizes must be the same, that is, $m = n$.

$\therefore$ A must be a square matrix.

E20) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a square matrix over a field F.

Then $A^t = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$

$$\therefore A + A^t = \begin{bmatrix} a+a & b+c \\ c+b & d+d \end{bmatrix} = \begin{bmatrix} 2a & b+c \\ b+c & 2d \end{bmatrix}, \text{ and}$$

$$A - A^t = \begin{bmatrix} a-a & b-c \\ c-b & d-d \end{bmatrix} = \begin{bmatrix} 0 & b-c \\ -(b-c) & 0 \end{bmatrix}.$$

You can check that $(A+A^t)^t = A + A^t$ and $(A - A^t)^t = -(A - A^t)$.

$\therefore A+A^t$ is symmetric and $A - A^t$ is skew-symmetric.

E21) $\begin{bmatrix} -i & 2 \\ 2 & -i \end{bmatrix}$.

E22) The size of $\bar{A}^t$ is the same as the size of A^t. $\therefore A = \bar{A}^t$ implies that the sizes of A and A^t are the same. $\therefore$ A is a square matrix.

E23) $I: R^n \to R^n : I(x_1,...,x_n) = (x_1,...., x_n)$.

Then, for any basis $B = \{e_1,..., e_n\}$ of R^n, $I(e_i) = e_i$.

$$\therefore [I]_B = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I_n$$

E24) Since A is upper triangular, all its elements below the diagonal are zero. Again, since $A = A^t$, a lower triangular matrix, all the entries of A above the diagonal are zero. $\therefore$, all the off-diagonal entries of A are zero. $\therefore$ A is a diagonal matrix.

E25) Let A be a skew-symmetric matrix. Then $A = -A^t$. Therefore,

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} -a_{11} & -a_{21} & \dots & -a_{n1} \\ -a_{12} & -a_{22} & \dots & -a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{1n} & -a_{2n} & \dots & -a_{nn} \end{bmatrix}$$

$\therefore$, for any $i = 1,..., n$, $a_{ii} = -a_{ii} \Rightarrow a_{ii} = 0 \Rightarrow a_{ij} = 0$.

The converse is not true. For example, the diagonal entries of $\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$ are zero, but this matrix is not skew-symmetric.

E26) $[1 \times 1 + 0 \times 2 + 0 \times 3] = [1]$

E27) $AB = \begin{bmatrix} 1 \times 1 + 1 \times 1 & 1 \times 0 + 1 \times 1 \\ 0 \times 1 + 1 \times 1 & 0 \times 0 + 1 \times 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$

$BA = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$

E 28) C + D is not defined.

CD is a 2×2 matrix and DC is a 3×3 matrix. $\therefore CD \neq DC$.

$CD = \begin{bmatrix} 1 \times 0 + 1 \times 1 + 0 \times 0 & 1 \times 1 + 1 \times 1 + 0 \times 0 \\ 0 \times 0 + 1 \times 1 + 0 \times 0 & 0 \times 1 + 1 \times 1 + 0 \times 0 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$

$$DC = \begin{bmatrix} 0 & 1 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 \times 1 + 1 \times 0 & 0 \times 1 + 1 \times 1 & 0 \times 0 + 1 \times 0 \\ 1 \times 1 + 1 \times 0 & 1 \times 1 + 1 \times 1 & 1 \times 0 + 1 \times 0 \\ 0 \times 1 + 0 \times 0 & 0 \times 1 + 0 \times 1 & 0 \times 0 + 0 \times 0 \end{bmatrix}.$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

E29) $A + B = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$. $\therefore (A + B)^2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$

Also $A^2 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$,

$B^2 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$,

$2 \cdot AB = 2\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$

$\therefore A^2 + 2AB + B^2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ 4 & 4 \end{bmatrix}$

$\therefore (A + B)^2 \neq A^2 + 2AB + B^2$

E30) $A^2 = \begin{bmatrix} -bd & b \\ d^2 b & db \end{bmatrix}\begin{bmatrix} -bd & b \\ d^2 b & db \end{bmatrix} = \begin{bmatrix} b^2 d^2 + d^2 b^2 & -b^2 d + db^2 \\ -b^2 d^3 + d^3 b^2 & d^2 b^2 + d^2 b^2 \end{bmatrix} = \begin{bmatrix} 2d^2 b^2 \\ 0 \end{bmatrix}$

E31) $\begin{bmatrix}1&0&0\\0&2&0\\0&0&3\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=\begin{bmatrix}x\\2y\\3z\end{bmatrix}$

$[x\ y\ z]\begin{bmatrix}1&0&0\\0&2&0\\0&0&3\end{bmatrix} = [x\quad 2y\quad 3z]$

E32) We take $A=\begin{bmatrix}1&2\\0&0\\3&1\end{bmatrix}$, $B=\begin{bmatrix}1&2&3&0\\4&5&1&1\end{bmatrix}$ Then

$AB=\begin{bmatrix}9&12&5&2\\0&0&0&0\\7&11&10&1\end{bmatrix}$, You can see that the 2nd row of AB is zero.

E33) $[S]_B=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\end{bmatrix}$, $[T]_B=\begin{bmatrix}1&0&0\\0&0&0\\0&1&0\end{bmatrix}$

$\therefore [S]_B[T]_B=\begin{bmatrix}0&0&0\\1&0&0\\0&0&0\end{bmatrix}$

Also, $[SoT]_B=\begin{bmatrix}0&0&0\\1&0&0\\0&0&0\end{bmatrix}=[S]_B[T]_B$

E34) $(A+B)^2=(A+B)(A+B)=A(A+B)+B(A+B)$ (by distributivity)

$=A^2+AB+BA+B^2$ (by distributivity)

E35) $AB=\begin{bmatrix}-1&-8&-10\\1&-2&-5\\9&22&15\end{bmatrix}$ $\therefore 2(AB)=\begin{bmatrix}-2&-16&-20\\2&-4&-10\\18&44&30\end{bmatrix}$

On the other hand, $(2A)B=\begin{bmatrix}4&-2\\2&0\\-6&8\end{bmatrix}\begin{bmatrix}1&-2&-5\\3&4&0\end{bmatrix}$

$=\begin{bmatrix}-2&-16&-20\\2&-4&-10\\18&44&30\end{bmatrix}$

$\therefore 2(AB)=(2A)B$

E36) $AB=\begin{bmatrix}0&-7&-3\\-11&-4&6\\0&0&0\end{bmatrix}$ $\therefore (AB)^t=\begin{bmatrix}0&-11&0\\-7&-4&0\\-3&6&0\end{bmatrix}$

Also, $B^tA^t=\begin{bmatrix}1&2&4\\-4&-1&0\\0&3&-2\end{bmatrix}\begin{bmatrix}2&1&0\\-1&0&0\\0&-3&0\end{bmatrix}=\begin{bmatrix}0&-11&0\\-7&-4&0\\-3&6&0\end{bmatrix}=(AB)^t$

E37) First, suppose AB is symmetric. Then $AB=(AB)^t=B^tA^t=BA$, since A and B are symmetric.

Conversely, suppose $AB=BA$. Then

$(AB)^t=B^tA^t=BA=AB$, so that AB is symmetric.

E38) Let $A=\text{diag}(d_1,\dots,d_n)$, $B=\text{diag}(e_1,\dots,e_n)$. Then

$$AB=\begin{bmatrix}d_1&0&0&\dots&0\\0&d_2&0&\dots&0\\ \vdots&\vdots&\vdots&&\vdots\\0&0&\dots&\dots&d_n\end{bmatrix}\begin{bmatrix}e_1&0&0&\cdots&0\\0&e_2&0&\cdots&0\\ \vdots&\vdots&\ddots&&\vdots\\0&0&\dots&\dots&e_n\end{bmatrix}$$

$$= \begin{bmatrix} d_1e_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2e_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3e_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_ne_n \end{bmatrix}$$

$= \text{diag}\,(d_1e_1, d_2e_2, \ldots, d_ne_n)$.

E39) Suppose it is invertible. Then $\exists\ A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ such that

$$\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$$

This gives us $A = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$ which is the same as the given matrix. This shows that the given matrix is invertible and, in fact, $\begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$

E40) Firstly, θ is a well defined map. Secondly, check that $\theta(v_1 + v_2) = \theta(v_1) + \theta(v_2)$, and $\theta(\alpha v) = \alpha\theta(v)$ for $v, v_1, v_2 \in V$ and $\alpha \in F$. Thirdly, show that $\theta(v) = 0 \Rightarrow v = 0$, that is θ is 1 – 1. Then, by Unit 5 (Theorem 10), you have shown that θ is an isomorphism.

E41) We will show that its columns are linearly independent over Q. Now, if $x, y, z \in$ Q such that

$x\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} + z\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, we get the equations

$2x + z = 0$

$z = 0$

$3y = 0$

On solving them we get $x = 0, y = 0, z = 0$.

$\therefore$ the given matrix is linearly independent.

E42) Let $B = \{e_1, e_2, e_3\}$ $B' = \{f_1, f_2, f_3\}$. Then

$f_1 = 0e_1 + 1e_2 + 0e_3 = e_2$

$f_2 = e_1 + e_2$

$f_3 = e_1 + 3e_3$

$\therefore B' = \{e_2, e_1 + e_2, e_1 + 3e_3\}$.

UNIT 8 MATRICES - II

Structure

8.1 INTRODUCTION

In Unit 7 we introduced you to a matrix and showed you how a system of linear equations can give us a matrix. An important reason for which linear algebra arose is the theory of simultaneous linear equations. A system of simultaneous linear equations can be translated into a matrix equation, and solved by using matrices.

The study of the rank of a matrix is a natural forerunner to the theory of simultaneous linear equations. Because, it is in terms of rank that we can find out whether a simultaneous system of equations has a solution or not. In this unit we start by studying the rank of a matrix. Then we discuss row operations on a matrix and use them for obtaining the rank and inverse of a matrix. Finally, we apply this knowledge to determine the nature of solutions of a system of linear equations. The method of solving a system of linear equations that we give here is by "successive elimination of variables". It is also called the Gaussian elimination process.

With this unit we finish Block 2. In the next block we will discuss concepts that are intimately related to matrices.

Objectives

After reading this unit, you should be able to

- obtain the rank of a matrix;
- reduce a matrix to the echelon form;
- obtain the inverse of a matrix by row-reduction;
- solve a system of simultaneous linear equations by the method of successive elimination of variables.

8.2 RANK OF A MATRIX

Consider any $m \times n$ matrix A, over a field F. We can associate two vector spaces with it, in a very natural way. Let us see what they are. Let $A = [a_{ij}]$. A has m rows, say, $R_1, R_2, \ldots, R_m$, where $R_1 = (a_{11}, a_{12}, \ldots, a_{1n})$, $R_2 = (a_{21}, a_{22}, \ldots, a_{2n}), \ldots, R_m = (a_{m1}, a_{m2}, \ldots, a_{mn})$.

Thus, $R_i \in F^n \; \forall i$, and $A = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{bmatrix}$

The subspace of F^n generated by the row vectors $R_1, \ldots, R_m$ of A, is called the **row space** of A, and is denoted by RS (A).

Example 1: If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, does $(0, 0, 1) \in$ RS (A)?

Solution : The row space of A is the subspace of R^3 generated by (1, 0, 0) and (0, 1, 0). Therefore, RS (A) = {(a, b, 0)| a, b ∈ R}. Therefore (0, 0, 1) ∉ RS (A).

ρ is the Greek letter 'rho'

The dimension of the row space of A is called the **row rank** of A, and is denoted by ρ_r (A).

Thus, ρ_r (A) = maximum number of linearly independent rows of A.

In Example 1, ρ_r(A) = 2 = number of rows of A. But consider the next example.

Example 2: If $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}$, find ρ_r (A).

Solution: The row space of A is the subspace of R^2 generated by (1,0), (0,1) and (2,0). But (2,0) already lies in the vector space generated by (1,0) and (0,1), since (2,0) = 2 (1,0). Therefore, the row space of A is generated by the linearly independent vectors (1, 0) and (0,1). Thus, ρ_r (A) = 2.

So, in Example 2, ρ_r (A) < number of rows of A.

In general, for any m × n matrix A, RS (A) is generated by m vectors. Therefore, $\rho_r (A) \le m$. Also, RS (A) is a subspace of F^n and $\dim_F F^n = n$. Therefore, $\rho_r (A) \le n$.

min (m, n) denotes 'the minimum of the numbers of m and n'.

Thus, **for any m × n matrix A,** $0 \le \rho_r (A) \le \min (m, n)$.

E E1) Show that $A = 0 \Leftrightarrow \rho_r (A) = 0$

Just as we have defined the row space of A, we can define the column space of A. Each column of A is an m-tuple, and hence belongs to F^m. We denote the columns of A by $C_1,\ldots,C_n$. The subspace of F^m generated by $\{C_1,\ldots,C_n\}$ is called the **column** space of A and is denoted by CS (A). The dimension of CS (A) is called the **column rank** of A, and is denoted of ρ_c (A). Again, since CS (A) is generated by n vectors and is a subspace of F^m, we get $0 \le \rho_c (A) \le \min (m, n)$.

E E2) Obtain the column rank and row rank of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}$

In E2 you may have noticed that the row and column ranks of A are equal. In fact, in Theorem 1, we prove that $\rho_r (A) = \rho_c (A)$, for any matrix A. But first, we prove a lemma.

Lemma 1: Let A, B be two matrices over F such that AB is defined. Then

a) CS (AB) ⊆ CS (A),

b) RS (AB) ⊆ RS (B).

Thus, $\rho_c (AB) \le \rho_c (A)$, $\rho_r (AB) \le \rho_r (B)$.

Proof: (a) Suppose $A = [a_{ij}]$ is an m × n matrix and $B = [b_{ij}]$ is an n × p matrix. Then, from Sec. 7.5, you know that the jth column of C = AB will be

$$\begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^{n} a_{1k} b_{kj} \\ \sum_{k=1}^{n} a_{2k} b_{kj} \\ \vdots \\ \sum_{k=1}^{n} a_{mk} b_{kj} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} b_{1j} + \ldots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} b_{nj}$$

$$= C_1 b_{1j} + \ldots + C_n b_{nj},$$

where $C_1, \ldots, C_n$ are the columns of A.

Thus, the columns of AB are linear combinations of the columns of A. Thus, the columns of $AB \in CS(A)$. So, $CS(AB) \subseteq CS(A)$.

Hence, $\rho_c(AB) \leq \rho_c(A)$.

b) By a similar argument as above, we get $RS(AB) \subseteq RS(B)$, and so, $\rho_r(AB) \leq \rho_r(B)$.

E E3) Prove (b) of Lemma 1.

We will now use Lemma 1 for proving the following theorem.

Theorem 1: $\rho_c(A) = \rho_r(A)$, for any matrix A over F.

Proof: Let $A \in \mathbf{M}_{m \times n}(\mathbf{F})$ Suppose $\rho_r(A) = r$ and $\rho_c(A) = t$.

Now, $RS(A) = [\{R_1, R_2, \ldots, R_m\}]$, where $R_1, R_2, \ldots, R_m$ are the rows of A. Let $\{e_1, e_2, \ldots, e_r\}$ be a basis of RS(A). Then R_i is a linear combination of $e_1, \ldots, e_r$, for each $i = 1, \ldots, m$. Let

$R_i = \sum_{j=1}^{r} b_{ij} e_j$, $i = 1, 2, \ldots, m$, where $b_{ij} \in F$ for $1 \leq i \leq m$, $1 \leq j \leq r$.

We can write these equations in matrix form as

$$\begin{bmatrix} R_1 \\ \vdots \\ R_m \end{bmatrix} = \begin{bmatrix} b_{11} \ldots b_{1r} \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ b_{m1} \ldots b_{mr} \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_r \end{bmatrix}.$$

So, $A = BE$, where $B = [b_{ij}]$ is an $m \times r$ matrix and E is the $r \times n$ matrix with rows $e_1, e_2, \ldots, e_r$. (Remember, $e_i \in F^n$, for each $i = 1, \ldots, r$.)

So, $t = \rho_c(A) = \rho_c(BE) \leq \rho_c(B)$, by Lemma 1,

$\leq \min(m, r)$

$\leq r$

Thus, $t \leq r$.

Just as we got A = BE above, we get $A = [f_1, \ldots, f_t]D$, where $\{f_1, \ldots, f_t\}$ is a basis of the column space of A and D is a $t \times n$ matrix. Thus, $r = \rho_r(A) \leq \rho_r(D) \leq t$, by Lemma 1.

So we get $r \leq t$ and $t \leq r$. This gives us $r = t$.

Theorem 1 allows us to make the following definition.

Definition: The integer $\rho_c(A)$ $(= \rho_r(A))$ is called the **rank** of A, and is denoted by $\rho(A)$.

You will see that Theorem 1 is very helpful if we want to prove any fact about $\rho(A)$. If it is easier to deal with the rows of A we can prove the fact for $\rho_r(A)$. Similarly, if it is easier to deal with the columns of A, we can prove the fact for $\rho_c(A)$. While proving Theorem 3 we have used this facility that Theorem 1 gives us.

Use Theorem 1 to solve the following exercises.

E E4) If A,B are two matrices such that AB is defined then show that $\rho(AB) \le \min(\rho(A), \rho(B))$.

E E5) Suppose $C \neq 0 \in M_{m \times 1}(F)$, and $R \neq 0 \in M_{1 \times n}(F)$, then show that the rank of the $m \times n$ matrix CR is 1. (Hint: Use E4).

Does the term 'rank' seem familiar to you? Do you remember studying about the rank of a linear transformation in Unit 5? We will now see if the rank of a linear transformation is related to the rank of its matrix. The following theorem brings forth the precise relationship. (Go through Sec. 5.3 before going further.)

Theorem 2: Let U,V be vector spaces over **F** of dimensions n and m, respectively. Let B_1 be a basis of U and B_2 be a basis of V. Let $T \in L(U,V)$.

Then $R(T) \simeq CS([T]_{B_1,B_2})$.

Thus, rank (T) = rank of $[T]_{B_1,B_2}$.

Proof: Let $B_1 = \{e_1, e_2, \ldots, e_n\}$ and $B_2 = \{f_1, f_2, \ldots, f_m\}$. As in the proof of Theorem 7 of Unit 7, $\theta: V \to M_{m \times 1}(F)$: $\theta(v)$ = coordinate vector of v with respect to the basis B_2, is an isomorphism.

Now, $R(T) = [\{T(e_1), T(e_2), \ldots, T(e_n)\}]$. Let $A = [T]_{B_1,B_2}$ have $C_1, C_2, \ldots, C_n$ as its columns. Then $CS(A) = [\{C_1, C_2, \ldots, C_n\}]$. Also, $\theta(T(e_i)) = C_i \; \forall i = 1, \ldots, n$.

Thus, $\theta: R(T) \to CS(A)$ is an isomorphism. $\therefore R(T) \simeq CS(A)$.

In particular, $\dim R(T) = \dim CS(A) = \rho(A)$.

That is, rank $(T) = \rho(A)$.

Theorem 2 leads us to the following corollary. It says that pre-multiplying or post-multiplying a matrix by invertible matrices does not alter its rank.

Corollary 1: Let A be an $m \times n$ matrix. Let P,Q be $m \times m$ and $n \times n$ invertible matrices, respectively.

Then $\rho(PAQ) = \rho(A)$.

Proof: Let $T \in L(U,V)$ be such that $[T]_{B_1,B_2} = A$. We are given matrices Q and P^{-1}. Therefore, by Theorem 8 of Unit 7, $\exists$ bases B_1' and B_2' of U and V, respectively, such that $Q = M_{B_1}^{B_1'}$ and $P^{-1} = M_{B_2}^{B_2'}$

Then, by Theorem 10 of Unit 7,

$[T]_{B_1', B_2'} = M_{B_2'}^{B_2} [T]_{B_1, B_2} M_{B_1}^{B_1'} = PAQ$

In other words, we can change the bases suitably so that the matrix of T with respect to the new bases is PAQ.

So, by Theorem 2, $\rho(PAQ)$ = rank (T) = $\rho(A)$. Thus, $\rho(PAQ) = \rho(A)$.

E6) Take $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \end{bmatrix}$, $P = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, $Q = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ Obtain PAQ and show that $\rho(PAQ) = \rho(A)$.

Now we state and prove another corollary to Theorem 2. This corollary is useful because it transforms any matrix into a very simple matrix, namely, a matrix whose entries are 1 and 0 only.

Corollary 2: Let A be an $m \times n$ matrix with rank r. Then $\exists$ invertible matrices P and Q such that $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

Proof: Let $T \in L(U,V)$ be such that $[T]_{B_1, B_2} = A$. Since $\rho(A) = r$, rank (T) = r. $\therefore$ nullity (T) = n − r (Unit 5, Theorem 5).

Let $\{u_1, u_2, \ldots, u_{n-r}\}$ be a basis of Ker T. We extend this to form the basis

$B_1' = \{u_1, u_2, \ldots, u_{n-r}, u_{n-r+1}, \ldots, u_n\}$ of U. Then $\{T(u_{n-r+1}), \ldots, T(u_n)\}$ is a basis of R(T) (See Unit 5, proof of Theorem 5). Extend this set to form a basis B_2' of V, say $B_2' = \{T(u_{n-r+1}), \ldots, T(u_n), v_1, \ldots, v_{m-r}\}$. Let us reorder the elements of B_1' and write it as $B_1' = \{u_{n-r+1}, \ldots, u_n, u_1, \ldots, u_{n-r}\}$.

Then, by definition, $[T]_{B_1', B_2'} = \begin{bmatrix} I_r & 0_{r \times (n-r)} \\ 0_{(m-r) \times r} & 0_{(m-r) \times (n-r)} \end{bmatrix}$

where $0_{s \times t}$ denotes the zero matrix of size s×t. (Remember that $u_1, \ldots, u_{n-r} \in$ Ker T.)

Hence, $PAQ = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$

where $Q = M_{B_1}^{B_1'}$ and $P = M_{B_2'}^{B_2}$, by Theorem 10 of Unit 7.

Note: $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ is called the **normal form** of the matrix A.

Consider the following example, which is the converse of E5.

Example 3: A is an $m \times n$ matrix of rank 1, show that $\exists$ $C \neq 0$ in $M_{m \times 1}$ (F) and $R \neq 0$ in $M_{1 \times n}$ (F) such that A = CR.

Solution: By Corollary 2 (above), $\exists$ P, Q such that

$$PAQ = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}, \text{ since } \rho(A) = 1.$$

$$= \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} [1\ 0 \ldots 0]$$

$$\therefore A = P^{-1}(PAQ)Q^{-1} = P^{-1}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}[1\ 0 \ldots 0]Q^{-1} = CR,$$

where $C = P^{-1}\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \neq 0$, $R = [1\ 0 \ldots 0]Q^{-1} \neq 0$

E7) What is the normal form of diag (1, 2, 3)?

The solution of E7 is a particular case of the general phenomenon: the normal form of an $n \times n$ invertible matrix is I_n.

Let us now look at some ways of transforming a matrix by playing around with its rows. The idea is to get more and more entries of the matrix to be zero. This will help us in solving systems of linear equations.

8.3 ELEMENTARY OPERATIONS

Consider the following set of 2 equations in 3 unknowns x, y and z:

$x + y + z = 1$

$2x + 3z = 0$

How can you express this system of equations in matrix form?
One way is

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

In general, if a system of m linear equations in n variables, $x_1, \ldots, x_n$, is

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$$

$$\vdots \qquad \vdots \qquad \qquad \vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m$$

where $a_{ij}, b_i \in F \ \forall\ i = 1, \ldots, m$ and $j = 1, \ldots, n$, then this can be expressed as
$AX = B$

where $A = [a_{ij}]_{m \times n}$, $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$

In this section we will study methods of changing the matrix A to a very simple form so that we can obtain an immediate solution to the system of linear equations $AX = B$. For this purpose, we will always be multiplying A on the left or the right by a suitable matrix. In effect, we will be applying elementary row or column operations on A.

8.3.1 Elementary Operations on a Matrix

Let A be an $m \times n$ matrix. As usual, we denote its rows by $R_1, \ldots, R_m$, and columns by $C_1, \ldots, C_n$. We call the following operations **elementary row operations:**

1) Interchanging R_i and R_j, for $i \neq j$.

2) Multiplying R_i by some $a \in \mathbf{F}$, $a \neq 0$.

3) Adding aR_j to R_i, where $i \neq j$ and $a \in \mathbf{F}$.

We denote the operation (1) by R_{ij}, (2) by $R_i(a)$, (3) by $R_{ij}(a)$.

For example, if $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \end{bmatrix}$,

then $R_{12}(A) = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$ (interchanging the two rows).

Also $R_2(3)(A) = \begin{bmatrix} 1 & 2 & 3 \\ 0 \times 3 & 1 \times 3 & 2 \times 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \end{bmatrix}$,

and $R_{12}(2)(A) = \begin{bmatrix} 1 + 0 \times 2 & 2 + 1 \times 2 & 3 + 2 \times 2 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 7 \\ 0 & 1 & 2 \end{bmatrix}$

E8) If $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, what is

a) $R_{21}(A)$ b) $R_{32} \circ R_{21}(A)$ c) $R_{13}(-1)(A)$?

Just as we defined the row operations, we can define the three column operations as follows:

1) Interchanging C_i and C_j for $i \neq j$, denoted by C_{ij}.

2) Multiplying C_i by $a \in \mathbf{F}$, $a \neq 0$, denoted by $C_i(a)$.

3) Adding aC_j to C_i, where $a \in \mathbf{F}$, denoted by $C_{ij}(a)$.

For example, if $A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$,

then $C_{21}(10)(A) = \begin{bmatrix} 1 & 13 \\ 2 & 24 \end{bmatrix}$

and $C_{12}(10)(A) = \begin{bmatrix} 31 & 3 \\ 42 & 4 \end{bmatrix}$

We will now prove a theorem which we will use in Sec. 8.3.2 for obtaining the rank of a matrix easily.

Theorem 3: Elementary operations on a matrix do not alter its rank.

Proof: The way we will prove the statement is to show that the row space remains unchanged under row operations and the column space remains unchanged under column operations. This means that the row rank and the column rank remain unchanged. This immediately shows, by Theorem 1, that the rank of the matrix remains unchanged.

Now, let us show that the row space remains unaltered. Let $R_1, \ldots, R_m$ be the rows of a matrix A. Then the row space of A is generated by $\{R_1, \ldots, R_i, \ldots, R_j, \ldots, R_m\}$. On applying R_{ij} to A, the rows of A remain the same. Only their order gets changed. Therefore, the row space of $R_{ij}(A)$ is the same as the row space of A.

If we apply $R_i(a)$, for $a \in \mathbf{F}$, $a \neq 0$, then any linear combination of $R_1, \ldots, R_m$ is $a_1R_1 + \ldots + a_mR_m$ $= a_1R_1 + \ldots + \frac{a_i}{a} aR_i + \ldots + a_mR_m$, which is a linear combination of $R_1, \ldots, aR_i, \ldots, R_m$.

Thus, $[\{R_1, \ldots, R_i, \ldots, R_m\}] = [\{R_1, \ldots, aR_i, \ldots, R_m\}]$. That is, the row space of A is the same as the row space of $R_i(a)(A)$.

If we apply $R_{ij}(a)$, for $a \in \mathbf{F}$, then any linear combination

$b_1R_1 + \ldots + b_iR_i + \ldots + b_jR_j + \ldots + b_mR_m = b_1R_1 + \ldots + b_i(R_i + aR_j) + \ldots + (b_j - b_ia)R_j + \ldots + b_mR_m$.

Thus, $[\{R_1, \ldots, R_m\}] = [\{R_1, \ldots, R_i + aR_j, \ldots, R_j, \ldots, R_m\}]$.

Hence, the row space of A remains unaltered under any elementary row operations.

We can similarly show that the column space remains unaltered under elementary column operations.

Elementary operations lead us to the following definition.

Definition: A matrix obtained by subjecting I_n to an elementary row or column operation is called an **elementary matrix**.

For example, $C_{12}(I_3) = C_{12}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is an elementary matrix.

Since there are six types of elementary operations, we get six types of elementary matrices, but not all of them are different.

E E9) Check that $R_{23}(I_4) = C_{23}(I_4)$, $R_2(2)(I_4) = C_2(2)(I_4)$ and $R_{12}(3)(I_4) = C_{21}(3)(I_4)$

In general, $R_{ij}(I_n) = C_{ij}(I_n)$, $R_i(a)(I_n) = C_i(a)(I_n)$ for $a \neq 0$, and $R_{ij}(a)(I_n) = C_{ji}(a)(I_n)$ for $i \neq j$ and $a \in \mathbf{F}$.

Thus, there are only three types of elementary matrices. We denote

$R_{ij}(I) = C_{ij}(I)$ by $\mathbf{E}_{ij}$,

$R_i(a)(I) = C_i(a)(I)$, (if $a \neq 0$) by $\mathbf{E}_i(a)$ and

$R_{ij}(a)(I) = C_{ji}(a)(I)$ by $\mathbf{E}_{ij}(a)$ for $i \neq j$, $a \in \mathbf{F}$.

E_{ij}, $E_i(a)$ and $E_{ij}(a)$ are called the elementary matrices corresponding to the pairs R_{ij} and C_{ij}, $R_i(a)$ and $C_i(a)$, $R_{ij}(a)$ and $C_{ji}(a)$, respectively.

Caution: $E_{ij}(a)$ corresponds to $C_{ji}(a)$, and not $C_{ij}(a)$.

Now, see what happens to the matrix

$A = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}$ if we multiply it on the left by

$E_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ We get

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & 2 \\ 2 & 1 & 0 \end{bmatrix} = R_{12}(A)$$

Similarly, $AE_{12} = C_{12}(A)$.

Again, consider $E_3(2)A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}$

$$= \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 4 & 2 & 0 \end{bmatrix} = R_3(2)(A)$$

Similarly, $AE_3(2) = C_3(2)(A)$

Finally, $E_{13}(5)A = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 10 & 6 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}$

$= R_{13}(5)(A)$

But, $AE_{13}(5) = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 0 \\ 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 3 & 0 & 15 \\ 2 & 1 & 10 \end{bmatrix}$

$= C_{31}(5)(A)$

What you have just seen are examples of a general phenomenon. We will now state this general result formally. (Its proof is slightly technical, and so, we skip it.)

Theorem 4: For any matrix A

a) $R_{ij}(A) = E_{ij}A$

b) $R_i(a)(A) = E_i(a)A$, for $a \neq 0$.

c) $R_{ij}(a)(A) = E_{ij}(a)A$

d) $C_{ij}(A) = AE_{ij}$

e) $C_i(a)(A) = AE_i(a)$, for $a \neq 0$

f) $C_{ij}(a)(A) = AE_{ji}(a)$

In (f) note the change of indices i and j.

An immediate corollary to this theorem shows that all the elementary matrices are invertible (see Sec. 7.6).

Corollary: An elementary matrix is invertible. In fact,

a) $E_{ij}E_{ij} = I$,

b) $E_i(a^{-1})E_i(a) = I$, for $a \neq 0$.

c) $E_{ij}(-a)E_{ij}(a) = I$.

Proof: We prove (a) only and leave the rest to you (see E10).

Now, from Theorem 4,

$E_{ij}E_{ij} = R_{ij}(E_{ij}) = R_{ij}(R_{ij}(I)) = I$, by definition of R_{ij}.

E10) Prove (b) and (c) of the corollary above.

The corollary tells us that the elementary matrices are **invertible and the inverse of an elementary matrix is also an elementary matrix of the same type.**

E 11) Actually multiply the two 4×4 matrices $E_{13}(-2)$ and $E_{13}(2)$ to get I_4.

And now we will introduce you to a very nice type of matrix, which any matrix can be transformed to by applying elementary operations.

8.3.2 Row-reduced Echelon Matrices

Consider the matrix

$$\begin{bmatrix} 1 & 0 & 9 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

In this matrix the three non-zero rows come before the zero row, and the first non-zero entry in each non-zero row is 1. Also, below this 1, are only zeros. This type of matrix has a special name, which we now give.

An echelon matrix is so-called because of the step-like structure of its non-zero rows.

Definition: An $m \times n$ matrix A is called a **row-reduced echelon matrix** if

a) the non-zero rows come before the zero rows,

b) in each non-zero row, the first non-zero entry is 1, and

c) the first non-zero entry in every non-zero row (after the first row) is to the right of the first non-zero entry in the preceding row.

Is $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \end{bmatrix}$ a row-reduced echelon matrix? Yes. It satisfies all the conditions of the definition. On the other hand, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ are not row-reduced echelon matrices, since they violate conditions (a), (b) and (c), respectively.

The matrix

$$\begin{bmatrix} 0 & 1 & 3 & 4 & 9 & 7 & 8 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 5 & 6 & 10 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 7 & 0 & 12 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is a 6×11 row-reduced echelon matrix. The dotted line in it is to indicate the step-like structure of the non-zero rows.

But, why bring in this type of a matrix? Well the following theorem gives us one good reason.

Theorem 5: The rank of a row-reduced echelon matrix is equal to the number of its non-zero rows.

Proof: Let $R_1, R_2, \ldots, R_r$ be the non-zero rows of an $m \times n$ row-reduced echelon matrix, E. Then RS(E) is generated by $R_1, \ldots, R_r$. We want to show that $R_1, \ldots, R_r$ are linearly independent. Suppose R_1 has its first non-zero entry in column k_1, R_2 in column k_2, and so on. Then, for any r scalars $c_1, \ldots, c_r$ such that $c_1R_1 + c_2R_2 + \ldots + c_rR_r = 0$, we immediately get

$$\begin{array}{ll} & \qquad\qquad k_1 \downarrow \qquad\qquad k_2 \downarrow \qquad\qquad k_r \downarrow \\ c_1 & [0, \ldots\ldots, 0, 1, *, \ldots\ldots\ldots, *, \ldots\ldots\ldots\ldots, *, \ldots\ldots, *] \\ + c_2 & [0, \ldots\ldots\ldots\ldots\ldots\ldots\ldots, 0, 1, *, \ldots\ldots\ldots, *, \ldots\ldots, *] \\ \vdots & \vdots \\ + c_r & [0, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, 0, 1, *, \ldots\ldots, *] \\ = & [0, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, 0] \end{array}$$

where * denotes various entries that we aren't bothering to calculate.

This equation gives us the following equations (when we equate the k_1th entries, the k_2th entries , ..., the k_rth entries on both sides of the equation):

$c_1 = 0, c_1(*) + c_2 = 0, ..., c_1(*) + c_2(*) + ... + c_{r-1}(*) + c_r = 0.$

On solving these equations we get

$c_1 = 0 = c_2 = ... = c_r$. $\therefore R_1, ..., R_r$ are linearly independent $\therefore \rho(E) = r$.

Not only is it easy to obtain the rank of an echelon matrix, one can also solve linear equations of the type AX = B more easily if A is in echelon form.

Now, here is some good news!

Every matrix can be transformed to the row echelon form by a series of elementary row operations. We say that the matrix is **reduced** to the row echelon form. Consider the following example.

Example 4: Let $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 2 & 4 & 1 & 10 & 2 \end{bmatrix}$

Reduce A to the row echelon form.

Solution: The first column of A is zero. The second column is non-zero. The (1,2)th element is 0. We want 1 at this position. We apply R_{12} to A and get

$$A_1 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 3 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 2 & 4 & 1 & 10 & 2 \end{bmatrix}$$

The (1,2)th entry has become 1. Now, we subtract multiples of the first row from other rows so that the (2,2)th, (3,2)th, (4,2)th and (5,2)th entries become zero. So we apply $R_{31}(-1)$, and $R_{51}(-2)$, and get

$$A_2 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 3 & 12 & 0 \end{bmatrix}$$

Now, beneath the entries of the first row we have zeros in the first 3 columns, and in the fourth column we find non-zero entries. We want 1 at the (2,4)th position, so we interchange the 2nd and 3rd rows. We get

$$A_3 = \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 4 & 1 \\ 0 & 0 & 0 & 3 & 12 & 0 \end{bmatrix}$$

We now subtract suitable multiples of the 2nd row from the 3rd, 4th and 5th rows so that the (3,4)th, (4,4)th and (5,4)th entries all become zero. $\therefore$

$$A_3 \overset{R_{42}(-1), R_{52}(-3)}{\sim} \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = A_4.$$

Now we have zeros below the entries of the 2nd row, except for the 6th column. The (3,6)th element is 1. We subtract suitable multiples of the 3rd row from the 4th and 5th rows so that the (4,6)th, (5,6)th elements become zero. $\therefore$,

$A \overset{R}{\sim} B$ means that on applying the operation R to A we get the matrix B.

$$A_1 \sim \begin{bmatrix} 0 & 1 & 2 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

And now we have achieved a row echelon matrix. Notice that we applied 7 elementary operations to A to obtain this matrix.

In general, we have the following theorem.

Theorem 6: Every matrix can be reduced to a row-reduced echelon matrix by a finite sequence of elementary row operations.

The proof of this result is just a repetition of the process that you went through in Example 4. For practice, we give you the following exercise.

E E12) Reduce the matrix $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$ to echelon form.

Theorem 6 leads us to the following definition.

Definition: If a matrix A is reduced to a row-reduced echelon matrix E by a finite sequence of elementary row operations then E is called a **row-reduced echelon form** (or, the row echelon form) of A. We now give a useful result that immediately follows from Theorems 3 and 5.

Theorem 7: Let E be a row-reduced echelon form of A. Then the rank of A = number of non-zero rows of E.

Proof: We obtain E from A by applying elementary operations. Therefore, by Theorem 3, $\rho(A) = \rho(E)$. Also, $\rho(E)$ = the number of non-zero rows of E, by Theorem 5.

Thus, we have proved the theorem.

Let us look at some examples to actually see how the echelon form of a matrix simplifies matters.

Example 5: Find $\rho(A)$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 6 \end{bmatrix},$$

by reducing it to its row-reduced echelon form.

Solution: $\begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & 6 \end{bmatrix} \overset{R_{21}(-1)}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 3 \end{bmatrix} \overset{R_2(1/3)}{\sim} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$

which is the desired row-reduced echelon form. This has 2 non-zero rows. Hence, $\rho(A) = 2$.

E E13) Obtain the row-reduced echelon form of the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 & 5 \\ 2 & 1 & 7 & 6 \\ 4 & 5 & 7 & 10 \end{bmatrix}$$

Hence determine the rank of the matrix.

By now you must have got used to obtaining row echelon forms. Let us discuss some ways of applying this reduction.

8.4 APPLICATIONS OF ROW-REDUCTION

In this section we shall see how to utilise row-reduction for obtaining the inverse of a matrix, and for solving a system of linear equations.

8.4.1 Inverse of a Matrix

In Theorem 4 you discovered that applying a row transformation to a matrix A is the same as multiplying it on the left by a suitable elementary matrix. Thus, applying a series of row transformations to A is the same as pre-multiplying A by a series of elementary matrices. This means that, after the nth row transformation we obtain the matrix $E_nE_{n-1} \ldots E_2E_1A$, where $E_1, E_2, \ldots, E_n$ are elementary matrices.

Now, how do we use this knowledge for obtaining the inverse of an invertible matrix? Suppose we have an n × n invertible matrix A. We know that A = IA, where $I = I_n$. Now, we apply a series of elementary row operations $E_1, \ldots, E_s$ to A so that A gets transformed to I_n. Thus,

$$I = E_sE_{s-1} \ldots E_2E_1A = E_sE_{s-1} \ldots E_2E_1 (IA)$$
$$= (E_sE_{s-1} \ldots E_2E_1I)A = BA$$

where $B = E_s \ldots E_1I$. Then, B is the inverse of A!

Note that we are reducing A to I, and not only to the row echelon form.

We illustrate this below.

Example 6: Determine if the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

is invertible. If it is invertible, find its inverse.

Solution: Can we transform A to I? If so, then A will be invertible.

$$\text{Now, } A = IA = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix}$$

To transform A we will be pre-multiplying it by elementary matrices. We will also be pre-multiplying IA by these matrices. Therefore, as A is transformed to I, the same transformations are done to I on the right hand side of the matrix equation given above. Now

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -5 \\ 0 & -5 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} A \text{ (applying } R_{21}(-2) \text{ and } R_{31}(-3) \text{ to A)}$$

$$\Rightarrow \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 5 \\ 0 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ 3 & 0 & -1 \end{bmatrix} A \quad \text{(applying } R_2(-1) \text{ and } R_3(-1))$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & -18 \end{bmatrix} = \begin{bmatrix} -3 & 2 & 0 \\ 2 & -1 & 0 \\ -7 & 5 & -1 \end{bmatrix} A \text{ (applying } R_{12}(-2) \text{ and } R_{32}(-5))$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & -7 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 2 & 0 \\ 2 & -1 & 0 \\ 7/18 & -5/18 & 1/18 \end{bmatrix} A \text{ (applying } R_3(-1/18))$$

$$\Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -5/18 & 1/18 & 7/18 \\ 1/18 & 7/18 & -5/18 \\ 7/18 & -5/18 & 1/18 \end{bmatrix} A \text{ (applying } R_{13}(7) \text{ and } R_{23}(-5))$$

Hence, A is invertible and its inverse is $B = 1/18 \begin{bmatrix} -5 & 1 & 7 \\ 1 & 7 & -5 \\ 7 & -5 & 1 \end{bmatrix}$

To make sure that we haven't made a careless mistake at any stage, check the answer by multiplying B with A. Your answer should be I.

E E14) Show that $\begin{bmatrix} 0 & 1 & 3 \\ 2 & 3 & 5 \\ 3 & 5 & 7 \end{bmatrix}$ is invertible. Find its inverse.

Let us now look at another application of row-reduction.

8.4.2 Solving a System of Linear Equations

Any system of m linear equations, in n unknowns $x_1, \ldots, x_n$, is

$$\begin{array}{lll} a_{11}x_1 + \ldots + & a_{1n}x_n = & b_1 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + \ldots + & a_{mn}x_n = & b_m, \end{array}$$

where all the a_{ij} and b_i are scalars.

This can be written in matrix form as

$$AX = B, \text{ where } A = [a_{ij}], X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

If B = 0, the system is called homogeneous. In this situation we are in a position to say how many linearly independent solutions the system of equations has.

heorem 8: The number of linearly independent solutions of the matrix equation $AX = 0$ n − r, where A is an $m \times n$ matrix and $r = \rho(A)$.

roof: In Unit 7 you studied that given the matrix A, we can obtain a linear transformation : $F^n \to F^m$ such that $[T]_{B,B'} = A$, where B and B′ are bases of F^n and F^m, respectively.

ow, $X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is a solution of $AX = 0$ if and only if it lies in Ker T (since $T(X) = AX$).

hus, the number of linearly independent solutions is dim Ker T = nullity (T) = n − nk (T) (Unit 5, Theorem 5.)

lso, rank (T) = $\rho(A)$ (Theorem 2)

hus, the number of linearly independent solutions is $n - \rho(A)$.

his theorem is very useful for finding out whether a homogeneous system has any non-ivial solutions or not.

:xample 7: Consider the system of 3 equations in 3 unknowns:

$x - 2y + z = 0$
$+ y \quad = 0$
$- 3z \quad = 0$

low many solutions does it have which are linearly independent over **R**?

iolution: Here our coefficient matrix, $A = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & -3 \end{bmatrix}$

hus, n = 3. We have to find r. For this, we apply the row-reduction method. We obtain

$A \sim \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$, which is in echelon form and has rank 3.

Thus, $\rho(A) = 3$.

Thus, the number of linearly independent solutions is $3 - 3 = 0$. This means that this system of equation has no non-zero solution.

n Example 7 the number of unknowns was equal to the number of equations, that is, n = m. What happens if n > m?

A system of m homogeneous equations in n unknowns has a non-zero solution if n > m. Why? Well, if n > m, then the rank of the coefficient matrix is less than or equal to m, and ence, less than n. So, n − r > 0. Therefore, at least one non-zero solution exists.

Note: If a system $AX = 0$ has one solution, X_0, then it has an infinite number of solutions of the form cX_0, $c \in F$. This is because $AX_0 = 0 \Rightarrow A(cX_0) = 0 \; \forall c \in F$.

E15) Give a set of linearly independent solutions for the system of equations

$x + 2y + 3z = 0$
$2x + 4y + z = 0$

Now consider the general equation AX = B, where A is an m × n matrix. We form the **augmented matrix** [A B]. This is an m × (n + 1) matrix whose last column is the matrix B. Here, we also include the case B = 0.

Interchanging equations, multiplying an equation by a non-zero scalar, and adding to any equation a scalar times some other equation does not alter the set of solutions of the system of equations. In other words, if we apply elementary row operations on [A B] then the solution set does not change.

The following result tells us under what conditions the system AX = B has a solution.

Theorem 9: The system of linear equations given by the matrix equation AX = B has a solution if $\rho(A) = \rho([A\ B])$.

Proof: AX = B represents the system

$$\begin{matrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{matrix}$$

This is the same as

$$\begin{matrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n - b_1 = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n - b_m = 0 \end{matrix}$$

which is represented by $[A\ B]\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$. Therefore, any solution of AX = B is also a solution of $[A\ B]\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$, and vice versa. By Theorem 8, this system has a solution if and only if $n + 1 > \rho([A\ B])$.

Now, if the equation $[A\ B]\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$ has a solution, say $\begin{bmatrix} c_1 \\ \vdots \\ c_{n+1} \end{bmatrix}$, then $c_1C_1 + c_2C_2 + \dots + c_nC_n = B$, where $C_1, \dots, C_n$ are the columns of A. That is, B is a linear combination of the C_i's. ∴ RS ([A B]) = RS (A). ∴ $\rho(A) = \rho([A\ B])$.

Conversely, if $\rho(A) = \rho([A\ B])$, then the number of linearly independent columns of A and [A B] are the same. Therefore, B must be a linear combination of the columns $C_1, \dots, C_n$ of A.

Let $B = a_1C_1 + \dots + a_nC_n$, $a_i \in F \ \forall\ i$.

Then a solution of AX = B is $X = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$

Thus, AX = B has a solution if and only if $\rho(A) = \rho([A\ B])$.

Remark: If A is invertible then the system AX = B has the unique solution $X = A^{-1}B$.

Now, once we know that the system given by AX = B is consistent, how do we find a solution? We utilise the method of successive (or Gaussian) elimination. This method is attributed to the famous German mathematician, Carl Friedrich Gauss (1777–1855) (see Fig. 1). Gauss was called the "prince of mathematicians" by his contemporaries. He did a great amount of work in pure mathematics as well as in the probability theory of errors, geodesy, mechanics, electro-magnetism and optics.

A system of equations is called consistent if it has a solution.

To apply the method of Gaussian elimination, we first reduce [A B] to its row echelon form, E. Then, we write out the equations $E\begin{bmatrix} X \\ -1 \end{bmatrix} = 0$ and solve them, which is simple.

Let us illustrate the method.

Example 8: Solve the following system by using the Gaussian elimination process.

$x + 2y + 3z = 1$

$2x + 4y + z = 2$

Solution: The given system is the same as

$$\begin{bmatrix}1 & 2 & 3 & 1\\ 2 & 4 & 1 & 2\end{bmatrix}\begin{bmatrix}x\\ y\\ z\\ -1\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

We first reduce the coefficient matrix to echelon form.

$$\begin{bmatrix}1 & 2 & 3 & 1\\ 2 & 4 & 1 & 2\end{bmatrix}\overset{R_{21}(-2)}{\sim}\begin{bmatrix}1 & 2 & 3 & 1\\ 0 & 0 & -5 & 0\end{bmatrix}\overset{R_2(-1/5)}{\sim}\begin{bmatrix}1 & 2 & 3 & 1\\ 0 & 0 & 1 & 0\end{bmatrix}$$

This gives us an equivalent system of equations, namely,

$x + 2y + 3z = 1$ and $z = 0$

These are, again, equivalent to $x = 1 - 2y$ and $z = 0$.

We get the solution in terms of a parameter. Put $y = \alpha$. Then $x = 1 - 2\alpha$, $y = \alpha$, $z = 0$ is a solution, for any scalar α. Thus, the solution set is $\{(1 - 2\alpha, \alpha, 0) \mid \alpha \in \mathbf{R}\}$.

Fig. 1: Carl Friedrich Gauss

Now let us look at an example where $B = 0$, that is, the system is homogeneous.

Example 9: Obtain a solution set of the simultaneous equations

$$x + 2y + 5t = 0$$
$$2x + y + 7z + 6t = 0$$
$$4x + 5y + 7z + 16t = 0$$

Solution: The matrix of coefficients is

$$A = \begin{bmatrix}1 & 2 & 0 & 5\\ 2 & 1 & 7 & 6\\ 4 & 5 & 7 & 16\end{bmatrix}$$

The given system is equivalent to $AX = 0$. A row-reduced echelon form of this matrix is

$$\begin{bmatrix}1 & 2 & 0 & 5\\ 0 & 1 & -7/3 & 4/3\\ 0 & 0 & 0 & 0\end{bmatrix}.$$

Then the given system is equivalent to

$$\left.\begin{aligned} x + y \qquad\qquad + 5t &= 0\\ y - (7/3)z + (4/3)t &= 0\end{aligned}\right\} \Rightarrow \begin{aligned} x &= (-14/3)z - (7/3)t\\ y &= (7/3)z - (4/3)t\end{aligned}$$

which is the solution in terms of z and t. Thus, the solution set of the given system of equations, in terms of two parameters α and β, is

$\{((-14/3)\alpha - (7/3)\beta, (7/3)\alpha - (4/3)\beta, \alpha, \beta) \mid \alpha, \beta \in \mathbf{R}\}$.

This is a two-dimensional vector subspace of $\mathbf{R}^4$ with basis

$\{(-14/3, 7/3, 1, 0), (-7/3, -4/3, 0, 1)\}$.

For practice we give you the following exercise.

E16) Use the Gaussian method to obtain solution sets of the following system of equations.

$$4x_1 - 3x_2 + x_3 - 7 = 0$$
$$x_1 - 2x_2 - 2x_3 - 3 = 0$$
$$3x_1 - x_2 + 2x_3 + 1 = 0$$

And now we are near the end of this unit.

8.5 SUMMARY

In this unit we covered the following points.

1) We defined the row rank, column rank and rank of a matrix, and showed that they are equal.
2) We proved that the rank of a linear transformation is equal to the rank of its matrix.
3) We defined the six elementary row and column operations.
4) We have shown you how to reduce a matrix to the row-reduced echelon form.
5) We have used the echelon form to obtain the inverse of a matrix.
6) We proved that the number of linearly independent solutions of a homogeneous system of equations given by the matrix equation AX = 0 is n − r, where r = rank of A, n = number of columns of A.
7) We proved that the system of linear equations given by the matrix equation AX = B is consistent if and only if $\rho(A) = \rho([A\ B])$.
8) We have shown you how to solve a system of linear equations by the process of successive elimination of variables, that is, the Gaussian method.

8.6 SOLUTIONS/ANSWERS

E1) A is the m × n zero matrix $\Leftrightarrow$ RS (A) = {0}. $\Leftrightarrow \rho_r(A) = 0$.

E2) The column space of A is the subspace of $\mathbf{R}^2$ generated by (1,0), (0,2), (1,1). Now $\dim_{\mathbf{R}} CS(A) \le \dim_{\mathbf{R}} \mathbf{R}^2 = 2$. Also (1,0) and (0,2), are linearly independent.
$\therefore$ {(1,0), (0, 2)} is a basis of CS(A), and $\rho_c(A) = 2$.

The row space of A is the subspace of $\mathbf{R}^3$ generated by (1,0,1) and (0,2,1). These vectors are linearly independent, and hence, form a basis of RS (A). $\therefore \rho_r(A) = 2$.

E3) The ith row of C = AB is

$[c_{i1}\ c_{i2} \dots c_{ip}]$

$$= \left[\sum_{k=1}^{n} a_{ik} b_{k1} \ \ \sum_{k=1}^{n} a_{ik} b_{k2} \dots \ \ \sum_{k=1}^{n} a_{ik} b_{kp} \right]$$

$= a_{i1}[b_{11}\ b_{12} \dots b_{1p}] + a_{i2}[b_{21}\ b_{22} \dots b_{2p}] + \dots + a_{in}[b_{n1}\ b_{n2} \dots b_{np}]$, a linear combination of the rows of B. $\therefore$ RS.(AB) $\subseteq$ RS(B) $\therefore \rho_r(AB) \le \rho_r(B)$.

E4) By Lemma 1, $\rho(AB) \le \rho_c(A) = \rho(A)$
Also $\rho(AB) \le \rho_r(B) = \rho(B)$.
$\therefore \rho(AB) \le \min(\rho(A), \rho(B))$.

E5) $\rho(CR) \le \min(\rho(C), \rho(R))$
But $\rho(C) \le \min(m, 1) = 1$. Also $C \ne 0$. $\therefore \rho(C) = 1$. $\therefore \rho(CR) \le 1$.

Now, if $C = \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$, $R = [b_1, \dots, b_n]$, then

$$CR = \begin{bmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & & \vdots \\ a_mb_1 & a_mb_2 & \dots & a_mb_n \end{bmatrix}$$

Since $C \neq 0$, $a_i \neq 0$, for some i. Similarly, $b_j \neq 0$, for some j. $\therefore a_ib_j \neq 0$. $\therefore CR \neq 0$.
$\therefore \rho(CR) \neq 0$. $\therefore \rho(CR) = 1$.

6) $PAQ = \begin{bmatrix} 0 & -2 & -2 \\ -3 & -4 & -3 \end{bmatrix}$. The rows of PAQ are linearly independent. $\therefore \rho(PAQ) = 2$. Also the rows of A are linearly independent. $\therefore \rho(A) = 2$. $\therefore \rho(PAQ) = \rho(A)$.

7) Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ Then $\rho(A) = 3$. $\therefore$ A's normal form is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

8) a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

b) $R_{12} \circ R_{21}(A) = R_{12}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\right) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

c) $\begin{bmatrix} 0 + 0 \times (-1) & 0 + 1 \times (-1) & 1 + 0 \times (-1) \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$

9) $R_{23}(I_4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = C_{23}(I_4)$

$R_2(2)(I_4) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = C_2(2)(I_4)$

$R_{12}(3)(I_4) = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = C_{21}(3)(I_4)$

10) $E_i(a^{-1}) E_i(a) = R_i(a^{-1})(E_i(a)) = R_i(a^{-1}) R_i(a)(I) = I$.
This proves (b).
$E_{ij}(-a) E_{ij}(a) = R_{ij}(-a)(E_{ij}(a)) = R_{ij}(-a)(R_{ij}(a)(I)) = I$, providing (c).

11) $E_{13}(-2) E_{13}(2) = \begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

12) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix} \underset{\sim}{R_{31}(-3)} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 0 \end{bmatrix} \underset{\sim}{R_{32}(5)} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

13) $\begin{bmatrix} 1 & 2 & 0 & 5 \\ 2 & 1 & 7 & 6 \\ 4 & 5 & 7 & 10 \end{bmatrix} \underset{\sim}{R_{21}(-2), R_{31}(-4)} \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & -3 & 7 & -4 \\ 0 & -3 & 7 & -10 \end{bmatrix}$

$R_2(-1/3) \begin{bmatrix} 1 & 2 & 0 & 5 \\ 0 & 1 & -7/13 & 4/3 \\ 0 & -3 & 7 & -10 \end{bmatrix}$

$$\underset{\sim}{R_{32}(3)}\begin{bmatrix}1 & 2 & 0 & 5\\ 0 & 1 & -7/3 & 4/3\\ 0 & 0 & 0 & -6\end{bmatrix} \qquad \underset{\sim}{R_3(-1/6)}\begin{bmatrix}1 & 2 & 0 & 5\\ 0 & 1 & -7/3 & 4/3\\ 0 & 0 & 0 & 1\end{bmatrix}$$

$\therefore \rho(A) = 3$

E14) $A = \begin{bmatrix}0 & 1 & 3\\ 2 & 3 & 5\\ 3 & 5 & 7\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}A.$

$$\Rightarrow \begin{bmatrix}2 & 3 & 5\\ 0 & 1 & 3\\ 3 & 5 & 7\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix}A \text{ (applying } R_{12})$$

$$\Rightarrow \begin{bmatrix}1 & 3/2 & 5/2\\ 0 & 1 & 3\\ 0 & 1/2 & -1/2\end{bmatrix} = \begin{bmatrix}0 & 1/2 & 0\\ 1 & 0 & 0\\ 0 & -3/2 & 1\end{bmatrix}A \text{ (applying } R_1(1/2), R_{31}(-3))$$

$$\Rightarrow \begin{bmatrix}1 & 0 & -2\\ 0 & 1 & 3\\ 0 & 0 & -2\end{bmatrix} = \begin{bmatrix}-3/2 & 1/2 & 0\\ 1 & 0 & 0\\ -1/2 & -3/2 & 1\end{bmatrix}A \text{ (applying } R_{12}(-3/2), R_{32}(-1/2))$$

$$\Rightarrow \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}-1 & 2 & -1\\ 1/4 & -9/4 & 3/2\\ 1/4 & 3/4 & -1/2\end{bmatrix}A \text{ (applying } R_3(-1/2), R_{23}(-3) \text{ and } R_{13}(2))$$

$\therefore A$ is invertible, and $A^{-1} = \begin{bmatrix}-1 & 2 & -1\\ 1/4 & -9/4 & 3/2\\ 1/4 & 3/4 & -1/2\end{bmatrix}$

E15) The given system is equivalent to

$$\begin{bmatrix}1 & 2 & 3\\ 2 & 4 & 1\end{bmatrix}\begin{bmatrix}x\\ y\\ z\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$$

Now, the rank of $\begin{bmatrix}1 & 2 & 3\\ 2 & 4 & 1\end{bmatrix}$ is 2. $\therefore$, the number of linearly independent solutions is $3 - 2 = 1$. $\therefore$, any non-zero solution will be a linearly independent solution. Now, the given equations are equivalent to

$x + 2y = -3z$ ..... (1)

$2x + 4y = -z$ ..... (2)

(−3) times Equation (2) added to Equation (1) gives $-5x - 10y = 0$.

$\therefore x = -2y$. Then (1) gives $z = 0$. Thus, a solution is $(-2, 1, 0)$. $\therefore$, a set of linearly independent solutions is $\{(-2, 1, 0)\}$.

Note that you can get several answers to this exercise. But any solution will be $\alpha(-2, 1, 0)$, for some $\alpha \in \mathbb{R}$.

E16) The augmented matrix is [A B]

$$= \begin{bmatrix}4 & -3 & 1 & 7\\ 1 & -2 & -2 & 3\\ 3 & -1 & 2 & -1\end{bmatrix}. \text{ Its row-reduced echelon form is}$$

$$\begin{bmatrix}1 & -2 & -2 & 3\\ 0 & 1 & 9/5 & -1\\ 0 & 0 & 1 & 5\end{bmatrix}$$

Thus, the given system of equations is equivalent to

$$x_1 - 2x_2 - 2x_3 = 3$$
$$x_2 + (9/5)x_3 = -1$$
$$x_3 = 5.$$

We can solve this system to get the unique solution $x_1 = -7, x_2 = -10, x_3 = 5$.

MEDIA NOTE (MTE-2, Block 2)

Video Programme : Linear Transformations and Matrices

Content coordinator : Dr. Parvin Sinclair, School of Sciences, IGNOU.

Producer : Sunil Das, Communication Division, IGNOU.

Introduction

In this note we will give a brief overview of what we have covered in the programme. Twice, during the programme, we have suggested some exercises that you may like to do after you've finished seeing it. We will also list these exercises in this note.

The aim of this programme is to help you in getting a better understanding of some of the concepts that we have dealt with in Block 2.

Before watching the programme we expect you to have finished going through Block 2.

Programme Summary

In this programme we introduce you to some geometric transformations, namely, rotations, scalings, reflections, projections and translations. You will see that, except for translation, all these are linear transformations. Then we discuss the method of obtaining the matrix associated to a linear transformation with respect to a specified pair of ordered bases. Finally, we talk about the composition of transformations and how to multiply matrices.

During the programme we have suggested that you try the following exercises.

E 1) Check if the scaling $g : \mathbf{R}^2 \to \mathbf{R}^2 : g(x, y) = (2x, 3y)$ is a linear transformation or not.

E 2) Is the function $h : \mathbf{R}^2 \to \mathbf{R}^2 : h(x, y) = (y, x)$ linear?

E3) What is the matrix associated to the projection

$p : \mathbf{R}^3 \to \mathbf{R} : p(x, y, z) = x$

with respect to the bases $B_1 = \{e_1 = (1, 0, 0), e_3 = (0, 0, 1)\}$ and $B_2 = \{f_1 = 1\}$

Answers

E 1) The scaling g is linear since

$g[(x, y) + (x', y')] = g(x, y) + g(x', y')$, and

$g[\alpha(x, y)] = \alpha g(x, y) \ \forall \ \alpha \in \mathbf{R}$.

E 2) Yes, because h satisfies conditions LT1 and LT2.

E 3) $p(e_1) = 1, \ p(e_2) = 0, \ p(e_3) = 0$.

$\therefore [p]_{B_1, B_2} = [1 \quad 0 \quad 0]$

NOTES

UTTAR PRADESH
RAJARSHI TANDON OPEN UNIVERSITY

UGMM-02
LINEAR ALGEBRA

Block

3

EIGENVALUES AND EIGENVECTORS

BLOCK 3 EIGENVALUES AND EIGENVECTORS

This block consists of three units in which we first introduce you to the theory of determinants, and give its applications in solving systems of linear equations.

The theory of determinants was originated by Leibniz in 1693 while studying systems of simultaneous linear equations. The mathematician Jacobi was perhaps the most prolific contributor to the theory of determinants. In fact, there is a particular kind of determinant that is named Jacobian, after him. The mathematicians Cramer and Bezout used determinants extensively for solving systems of linear equations.

In Unit 9 we have given a self-contained treatment of determinants, including the standard properties of determinants. We have also given the formula for obtaining the inverse of a matrix, and have explained Cramer's Rule. We end this unit by discussing the determinant rank.

In Unit 10 we discuss eigenvalues and eigenvectors. Their use first appeared in the study of quadratic forms. (You will study such forms in the next block.) The concepts that you will study in this unit were developed by Arthur Cayley and others during the 1840s. What you will discover in the unit is the algebraic eigenvalue problem and methods of finding eigenvalues and linearly independent eigenvectors.

In Unit 11 we introduce you to the characteristic polynomial. We give a proof of the Cayley-Hamilton theorem and give its applications. We also discuss the minimal polynomial of a matrix and of a linear transformation.

If you are interested in knowing more about the material covered in this block, you can refer to the books listed in the course introduction. These books will be available at your study centre.

NOTATIONS AND SYMBOLS

$M_n(F)$	set of all $n \times n$ matrices over F
$V_n(F)$	$M_{n \times 1}(F)$
det (A) / $\|A\|$	determinant of the matrix A
$\prod_{i \in P} a_i$	the product of a_is such that i satisfies property P
det (T)	determinant of the linear operator T
Adj (A)	adjoint of the matrix A
Tr (A)	trace of the matrix A
W_λ	eigenspace corresponding to the eigenvalue λ

UNIT 9 DETERMINANTS

Structure

9.1 INTRODUCTION

In Unit 8 we discussed the successive elimination method for solving a system of linear equations. In this unit we introduce you to another method, which depends on the concept of a determinant function. Determinants were used by the German mathematician Leibniz (1646-1716) and the Swiss mathematician Cramer (1704-1752) to solve a system of linear equations. In 1771, the mathematician Vandermonde (1735-1796) gave the first systematic presentation of the theory of determinants.

There are several ways of developing the theory of determinants. In Section 9.2 we approach it in one way. In Section 9.3 you will study the properties of determinants and certain other basic facts about them. We go on to give their applications in solving a system of linear equations (Cramer's Rule) and obtaining the inverse of a matrix. We also define the determinant of a linear transformation. We end with discussing a method of obtaining the rank of a matrix.

Throughout this unit F will denote a field of characteristic zero (see Unit 1), $M_n(F)$ will denote the set of $n \times n$ matrices over F and $V_n(F)$ will denote the space of all $n \times 1$ matrices over F, that is,

$$V_n(F) = \left\{ X = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \middle| \, a_i \in F \right\}$$

The concept of a determinant must be understood properly because you will be using it again and again. Do spend more time on Section 9.2, if necessary. We also advise you to revise Block 2 before starting this unit.

Objectives

After completing this unit, you should be able to

- evaluate the determinant of a square matrix, using various properties of determinants;
- obtain the adjoint of a square matrix;
- compute the inverse of an invertible matrix, using its adjoint;
- apply Cramer's Rule to solve a system of linear equations;
- evaluate the determinant of a linear transformation;
- evaluate the rank of a matrix by using the concept of the determinant rank.

9.2 DEFINING DETERMINANTS

There are many ways of introducing and defining the determinant function from $M_n(F)$ to F. In this section we give one of them, the classical approach. This was given by the French mathematician Laplace (1749-1827), and is still very much in use.

We will define the determinant function det: $M_n(F) \to F$ by induction on n. That is, we will define it for n = 1, 2, 3, and then define it for any n, assuming the definition for n − 1.

When n = 1, for any $A \in M_1(F)$ we have $A = [a]$, for some $a \in F$. In this case we define $\det(A) = \det([a]) = a$.

For example, $\det([5]) = 5$ and $\det([-5]) = -5$.

When n = 2, for any $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \in M_2(F)$, we define $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

For example, $\det\left(\begin{bmatrix} 0 & 1 \\ -2 & 3 \end{bmatrix}\right) = 0 \times 3 - 1 \times (-2) = 2.$

When n = 3, for any $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in M_3(F)$, we define

det(A) using the definition for the case n = 2 as follows:

$$\det(A) = a_{11}\det\left(\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}\right) - a_{12}\det\left(\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}\right) + a_{13}\det\left(\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right).$$

That is, $\det(A) = (-1)^{1+1} a_{11}$ (det of the matrix left after deleting the row and column containing a_{11}) + $(-1)^{1+2} a_{12}$ (det of the matrix left after deleting the row and column containing a_{12}) + $(-1)^{1+3} a_{13}$ (det of the matrix left after deleting the row and column containing a_{13}).

Note that the power of (−1) that is attached to a_{1j} is 1 + j for j = 1, 2, 3.

We denote det (A) by |A| also. For example, the determinant of $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is denoted by $\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$

So, $\det(A) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$.

In fact, we could have calculated |A| from the second row also as follows:

$$|A| = (-1)^{2+1} a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + (-1)^{2+2} a_{22}\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + (-1)^{2+3} a_{23}\begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}.$$

Similarly, expanding by the third row, we get

$$|A| = (-1)^{3+1} a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} + (-1)^{3+2} a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + (-1)^{3+3} a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

All 3 ways of obtaining |A| lead to the same value.

Consider the following example.

Example 1 : Let

$A = \begin{bmatrix} 1 & 2 & 6 \\ 5 & 4 & 1 \\ 7 & 3 & 2 \end{bmatrix}$. Calculate |A|.

Solution: We want to obtain

$$|A| = \begin{vmatrix} 1 & 2 & 6 \\ 5 & 4 & 1 \\ 7 & 3 & 2 \end{vmatrix}$$

Let A_{ij} denote the matrix obtained by deleting the ith row and jth column of A.
Let us expand by the first row. Observe that

$$A_{11} = \begin{bmatrix} 4 & 1 \\ 3 & 2 \end{bmatrix}, A_{12} = \begin{bmatrix} 5 & 1 \\ 7 & 2 \end{bmatrix}, A_{13} = \begin{bmatrix} 5 & 4 \\ 7 & 3 \end{bmatrix}.$$

Thus,

$$|A_{11}| = \begin{vmatrix} 4 & 1 \\ 3 & 2 \end{vmatrix} = 4 \times 2 - 1 \times 3 = 5, |A_{12}| = \begin{vmatrix} 5 & 1 \\ 7 & 2 \end{vmatrix} = 5 \times 2 - 1 \times 7 = 3, |A_{13}| = \begin{vmatrix} 5 & 4 \\ 7 & 3 \end{vmatrix}$$
$$= 5 \times 3 - 4 \times 7 = -13.$$

Thus,

$$|A| = (-1)^{1+1} \times 1 \times |A_{11}| + (-1)^{1+2} \times 2 \times |A_{12}| + (-1)^{1+3} \times 6 \times |A_{13}| = 5 - 6 - 78$$
$$= -79.$$

E1) Now obtain $|A|$ of Example 1, by expanding by the second row, and the third row. Does the value of $|A|$ depend upon the row used for calculating it?

Now, let us see how this definition is extended to define det(A) for any $n \times n$ matrix A, $n \neq 1$.

When $A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \in M_n(F)$, we define det(A) by expanding from

the ith row as follows:

$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) + (-1)^{i+2} a_{i2} \det(A_{i2}) + \ldots + (-1)^{i+n} a_{in} \det(A_{in})$, where A_{ij} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and the jth column, and i is a fixed integer with $1 \leq i \leq n$.

We, thus, see that $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$, defines the determinant of an $n \times n$ matrix A in terms of the determinants of the $(n-1) \times (n-1)$ matrices A_{ij}, j = 1, 2,......, n.

Note: While calculating $|A|$, we prefer to expand along a row that has the maximum number of zeros. This cuts down the number of terms to be calculated.

The following example will help you to get used to calculating determinants.

Example 2: Let

$$A = \begin{bmatrix} -3 & -2 & 0 & 2 \\ 2 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & -3 & 1 \end{bmatrix}. \quad \text{Calculate } |A|.$$

Solution

$$|A| = \begin{vmatrix} -3 & -2 & 0 & 2 \\ 2 & 1 & 0 & -1 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & -3 & 1 \end{vmatrix}$$

The first three rows have one zero each. Let us exp...... ...ong the third row. Observe that $a_{32} = 0$. So we don't need to calculate $|A_{32}|$. Now,

$$A_{31} = \begin{bmatrix} -2 & 0 & 2 \\ 1 & 0 & -1 \\ 1 & -3 & 1 \end{bmatrix}, A_{33} = \begin{bmatrix} -3 & -2 & 2 \\ 2 & 1 & -1 \\ 2 & 1 & 1 \end{bmatrix}, A_{34} = \begin{bmatrix} -3 & -2 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & -3 \end{bmatrix}.$$

We will obtain $|A_{31}|$, $|A_{33}|$, and $|A_{34}|$ by expanding along the second, third and second row, respectively.

$$\therefore |A_{31}| = \begin{vmatrix} -2 & 0 & 2 \\ 1 & 0 & -1 \\ 1 & -3 & 1 \end{vmatrix}$$

$$= (-1)^{2+1}.1. \begin{vmatrix} 0 & 2 \\ -3 & 1 \end{vmatrix} + (-1)^{2+2}.0. \begin{vmatrix} -2 & 2 \\ 1 & 1 \end{vmatrix} + (-1)^{2+3}.(-1). \begin{vmatrix} -2 & 0 \\ 1 & -3 \end{vmatrix}$$

(expansion along the second row)

$= (-1).6 + 0 + (-1)(-1).6$
$= -6 + 6 = 0.$

$$|A_{33}| = \begin{vmatrix} -3 & -2 & 2 \\ 2 & 1 & -1 \\ 2 & 1 & 1 \end{vmatrix} = (-1)^{3+1}.2. \begin{vmatrix} -2 & 2 \\ 1 & -1 \end{vmatrix} + (-1)^{3+2}.1. \begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix}$$

$$+ (-1)^{3+3}.1. \begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix} \text{ (expansion along the third row)}$$

$= 1.2.0 + (-1).1.(-1) + 1.1.1$
$= 1 + 1 = 2.$

$$|A_{34}| = \begin{vmatrix} -3 & -2 & 0 \\ 2 & 1 & 0 \\ 2 & 1 & -3 \end{vmatrix} = (-1)^{2+1}.2. \begin{vmatrix} -2 & 0 \\ 1 & -3 \end{vmatrix} + (-1)^{2+2}.1. \begin{vmatrix} -3 & 0 \\ 2 & -3 \end{vmatrix}$$

$+(-1)^{2+3}\ .0.\begin{vmatrix} -3 & -2 \\ 2 & 1 \end{vmatrix}$ (expansion along the second row)

$= (-1)\ .2.6 + 1.1.9 + 0$
$= -12 + 9 = -3.$

Thus, the required determinant is given by

$|A| = a_{31}|A_{31}| - a_{32}|A_{32}| + a_{33}|A_{33}| - a_{34}|A_{34}|$
$= 1.0 - 0 + 1.2 - 2.(-3) = 8.$

E2) Calculate $|A^t|$, where A is the matrix in

a) Example 1,
b) Example 2.

At this point we mention that there are two other methods of obtaining determinants – via permutations and via multilinear forms. We will not be doing these methods here. For purposes of actual calculation of determinants the method that we have given is normally used. The other methods are used to prove various properties of determinants.

So far we have looked at determinants algebraically only. But there is a geometrical interpretation of determinants also, which we now give.

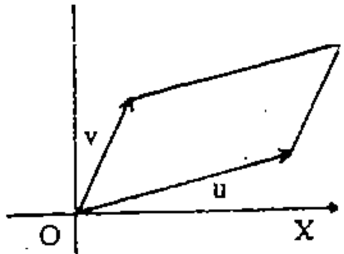


Fig. 1: The shaded area is det (u, v)

Determinant as area and volume: Let $u = (a_1, a_2)$ and $v = (b_1, b_2)$ be two vectors in R^2. Then, the magnitude of the area of the parallelogram spanned by u and v (see Fig. 1) is the absolute value of det (u, v) $= \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$.

In fact, what we have just said is true for any $n > 0$. Thus, if $u_1, u_2, \ldots, u_n$ are n vectors in

$\mathbf{R}^n$, then the absolute value of det $(u_1, u_2, \ldots, u_n)$ is the magnitude of the volume of the n-dimensional box spanned by $u_1, u_2, \ldots, u_n$.

Det $(C_1, C_2, \ldots, C_n)$ denotes det (A), where $A = (C_1, C_2, \ldots, C_n)$ is the matrix whose columns are $C_1, C_2, \ldots, C_n$.

Try this exercise now.

E E3) What is the magnitude of the volume of the box in $\mathbf{R}^3$ spanned by **i, j** and **k**?

Let us, now study some properties of the determinant function.

9.3 PROPERTIES OF DETERMINANTS

In this section we will state some properties of determinants, mostly without proof. We will take examples and check that these properties hold for them.

Now, for any $A \in M_n(F)$ we shall denote its columns by $C_1, C_2, \ldots, C_n$. Then we have the following 7 properties, P1–P7.

P1: If C_i' is an $n \times 1$ vector over **F**, then
det $(C_1, \ldots, C_{i-1}, C_i + C_i', C_{i+1}, \ldots, C_n)$
= det $(C_1, \ldots, C_{i-1}, C_i, C_{i+1}, \ldots, C_n)$ + det$(C_1, \ldots, C_{i-1}, C_i', C_{i+1}, \ldots, C_n)$.

P2: If $C_i = C_j$ for any $i \neq j$, then det $(C_1, C_2, \ldots, C_n) = 0$.

P3: If C_i and C_j are interchanged $(i \neq j)$ to form a new matrix B, then
det B = – det $(C_1, C_2, \ldots, C_n)$.

P4: For $\alpha \in$ **F**,
det $(C_1, \ldots, C_{i-1}, \alpha C_i, C_{i+1}, \ldots, C_n) = \alpha$ det $(C_1, C_2, \ldots, C_n)$.
Thus, det $(\alpha C_1, \alpha C_2, \ldots, \alpha C_n) = \alpha^n$ det $(C_1, \ldots, C_n)$.

Now, using P1, P2 and P4, we find that for $i \neq j$ and $\alpha \in$ **F**,
det $(C_1, \ldots, C_i + \alpha C_j, \ldots, C_j, \ldots, C_n)$ = det $(C_1, \ldots, C_i, \ldots, C_j, \ldots, C_n)$ + αdet $(C_1, \ldots, C_j, \ldots C_j, \ldots C_n)$.
= det $(C_1, C_2, \ldots, C_n)$.

Thus, we have
P5: For any $\alpha \in$ **F** and $i \neq j$, det $(C_1, \ldots, C_i + \alpha C_j, C_{i+1}, \ldots, C_n)$ = det $(C_1, C_2, \ldots, C_n)$.

Another property that we give is
P6: det(A) = det(A^t) $\forall A \in M_n(F)$. (In E2 you saw that this property was true for Examples 1 and 2. Its proof uses the permutation approach to determinants.)

Using P6, and the fact that det (A) can be obtained by expanding along any row, we get
P7: For $A \in M_n(F)$, we can obtain det(A) by expanding along any column. That is, for a fixed k,
$|A| = (-1)^{1+k} a_{1k} |A_{1k}| + (-1)^{2+k} a_{2k} |A_{2k}| + \ldots + (-1)^{n+k} a_{nk} |A_{nk}|$.

An important remark now.

Remark: Using P6, we can immediately say that P1–P5 are valid when columns are replaced by rows.

Using the notation of Unit 8, P3 says that

det $(R_{ij}(A))$ = – det(A) = det $(C_{ij}(A))$.

P4 says that
det $(R_i(\alpha)(A)) = \alpha^n$ det (A) = det $(C_i(\alpha)(A))$, $\forall \alpha \in$ **F**, and P5 says that
det $(R_{ij}(\alpha)(A))$ = det (A) = det $(C_{ij})(\alpha)(A)$, $\forall \alpha \in$ F.

We will now illustrate how useful the properties P1 – P7 are.

Example 3: Obtain det (A), where A is

a) $\begin{bmatrix} 1 & 6 & 0 \\ 2 & 7 & 2 \\ 1 & 6 & 0 \end{bmatrix}$ b) $\begin{bmatrix} 1 & 2 & -1 & -3 \\ 2 & 4 & 5 & 0 \\ 0 & 2 & -1 & -2 \\ -1 & 0 & 0 & 1 \end{bmatrix}$

Solution: a) Since the first and third rows of A (R_1 and R_3) coincide, $|A| = 0$, by P2 and P6.

b)

$$|A| = \begin{vmatrix} 1 & 2 & -1 & -3 \\ 2 & 4 & 5 & 0 \\ 0 & 2 & -1 & -2 \\ -1 & 0 & 0 & 1 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 2 & -1 & -3 \\ 2 & 4 & 5 & 0 \\ 0 & 2 & -1 & -2 \\ 0 & 2 & -1 & -2 \end{vmatrix}, \text{ by adding } R_1 \text{ to } R_4.$$

$= 0$, since $R_3 = R_4$.

Try the following exercise now.

E4) Calculate $\begin{vmatrix} 1 & 3 & 0 \\ 2 & 1 & 2 \\ 1 & 3 & 0 \end{vmatrix}$ and $\begin{vmatrix} 2 & 3 & 5 \\ 1 & 0 & 1 \\ 4 & 6 & 10 \end{vmatrix}$.

Now we give some examples of determinants that you may come across often.

Example 4: Let

$$A = \begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}, \text{ where } a, b \in \mathbf{R}.$$

Calculate $|A|$.

Solution:

$$|A| = \begin{vmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{vmatrix}$$

$$= \begin{vmatrix} a+3b & a+3b & a+3b & a+3b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{vmatrix}$$ (by adding the second, third and fourth rows to the first row, and applying P5)

$$= \begin{vmatrix} a+3b & 0 & 0 & 0 \\ b & a-b & 0 & 0 \\ b & 0 & a-b & 0 \\ b & 0 & 0 & -b \end{vmatrix}$$ (by subtracting the first column from every other column, and using P5)

$$= (a+3b)\begin{vmatrix} a-b & 0 & 0 \\ 0 & a-b & 0 \\ 0 & 0 & a-b \end{vmatrix}$$ (expanding along the first row)

$$= (a+3b)\ (a-b)^3$$

In Example 4 we have used an important, and easily proved fact, namely, **det (diag** $(\alpha_1, \alpha_2, \ldots, \alpha_n)) = \alpha_1 \alpha_2 \ldots \alpha_n, \alpha_i \in F \forall i$.

This is true because,

$$\begin{vmatrix} \alpha_1 & 0 & 0 \cdots & 0 \\ 0 & \alpha_2 & 0 \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 \cdots & \alpha_n \end{vmatrix} = \alpha_1 \alpha_2 \ldots \alpha_n \begin{vmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{vmatrix}, \text{by P4}$$

$$= \alpha_1 \alpha_2 \ldots \alpha_n |I_n|$$
$$= \alpha_1 \alpha_2 \ldots \alpha_n, \text{ since } |I_n| = 1.$$

$\prod_{i=1}^{n} a_i = a_1 a_2 \ldots a_n.$

Example 5: Show that

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ x_1^2 & x_2^2 & x_3^2 & x_4^2 \\ x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{vmatrix} = \prod_{i<j} (x_j - x_i), 1 \le i < j \le 4$$

(This is known as the Vandermonde's determinant of order 4)

Solution: The given determinant

$$= \begin{vmatrix} 1 & 0 & 0 & 0 \\ x_1 & x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ x_1^2 & x_2^2 - x_1^2 & x_3^2 - x_1^2 & x_4^2 - x_1^2 \\ x_1^3 & x_2^3 - x_1^3 & x_3^3 - x_1^3 & x_4^3 - x_1^3 \end{vmatrix}$$ (by subtracting the first column from every other column)

$$= \begin{vmatrix} x_2 - x_1 & x_3 - x_1 & x_4 - x_1 \\ (x_2 - x_1)(x_2 + x_1) & (x_3 - x_1)(x_3 + x_1) & (x_4 - x_1)(x_4 + x_1) \\ (x_2 - x_1)(x_2^2 + x_1^2 + x_2x_1) & (x_3 - x_1)(x_3^2 + x_1^2 + x_3x_1) & (x_4 - x_1)(x_4^2 + x_1^2 + x_4x_1) \end{vmatrix}$$

(by expanding along the first row and factorising the entries)

$$= (x_2-x_1)(x_3-x_1)(x_4-x_1) \begin{vmatrix} 1 & 1 & 1 \\ x_2 + x_1 & x_3 + x_1 & x_4 + x_1 \\ x_2^2 + x_1^2 + x_2x_1 & x_3^2 + x_1^2 + x_3x_1 & x_4^2 + x_1^2 + x_4x_1 \end{vmatrix}$$

(by taking out (x_2-x_1), (x_3-x_1), and (x_4-x_1) from Columns 1, 2 and 3 respectively).

$$= (x_2 - x_1)(x_3 - x_1)(x_4 - x_1) \begin{vmatrix} 1 & 0 & 0 \\ x_2 + x_1 & x_3 - x_2 & x_4 - x_2 \\ x_2^2 + x_1^2 + x_2x_1 & x_3^2 - x_2^2 + (x_3 - x_2)x_1 & x_4^2 - x_2^2 + (x_4 - x_2)x_1 \end{vmatrix}$$

(by subtracting the first column from the second and third columns)

$$=(x_2-x_1)(x_3-x_1)(x_4-x_1)\begin{vmatrix} x_3-x_2 & x_4-x_2 \\ (x_3-x_2)(x_3+x_2+x_1) & (x_4-x_2)(x_4+x_2+x_1)\end{vmatrix}$$

(expanding by the first row and factorising the entries)

$$= (x_2-x_1)(x_3-x_1)(x_4-x_1)(x_3-x_2)(x_4-x_2)\begin{vmatrix} 1 & 1 \\ x_3+x_2+x_1 & x_4+x_2+x_1\end{vmatrix}$$

$$= (x_2-x_1)(x_3-x_1)(x_4-x_1)(x_3-x_2)(x_4-x_2)(x_4-x_3)$$

$$= \prod_{i<j}(x_j - x_i),\ 1 \leqslant i, j \leqslant 4.$$

Try the following exercise now.

E5) What are $\begin{vmatrix} a & 0 & 0 \\ \alpha & b & 0 \\ \beta & \tau & c\end{vmatrix}$ and $\begin{vmatrix} a & d & e \\ 0 & b & f \\ 0 & 0 & c\end{vmatrix}$?

The answer of E4 is part of a general phenomenon, namely, the **determinant of an upper or lower triangular matrix is the product of its diagonal elements.**

The proof of this is immediate because,

$$\begin{vmatrix} a_{11} & * & \dots & * \\ 0 & a_{22} & \dots & * \\ 0 & 0 & & \cdot \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn}\end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & * & \dots & * \\ 0 & a_{33} & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn}\end{vmatrix} \text{ (expanding along } C_1)$$

$= \dots = a_{11}\, a_{22} \dots a_{nn}$, each time expanding along the first column.

In the Calculus course you must have come across $df/dt = f'(t)$, where f is a function of t. The next exercise involves this.

E6) Let us define the function $\theta(t)$ by

$$\theta(t) = \begin{vmatrix} f(t) & g(t) \\ f'(t) & g'(t)\end{vmatrix}$$

Show that $\theta'(t) = \begin{vmatrix} f(t) & g(t) \\ f''(t) & g''(t)\end{vmatrix}$

And now, let us study a method for obtaining the inverse of an invertible matrix.

9.4 INVERSE OF A MATRIX

In this section we first obtain the determinant of the product of two matrices and then define an adjoint of a matrix. Finally, we see the conditions under which a matrix is invertible, and, when it is invertible, we give its inverse in terms of its adjoint.

9.4.1 Product Formula

In Unit 7 you studied matrix multiplication. Let us see what happens to the determinant of a product of matrices.

Theorem 1: Let A and B be n × n matrices over **F**. Then det (AB) = det(A) det (B).

We will not do the proof here since it is slightly complicated. But let us verify Theorem 1 for some cases.

Example 6: Calculate |A|.|B| and |AB| when

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 10 & 9 \\ 0 & 3 & 8 \\ 0 & 0 & 5 \end{bmatrix}.$$

Solution: We want to verify Theorem 1 for our pair of matrices. Now, on expanding by the third row, we get |A| = 1.

Also, |B| = 30, which can be immediately seen since B is a triangular matrix.

$$\text{Since } AB = \begin{bmatrix} 2 & 10 & 19 \\ 6 & 33 & 35 \\ 0 & 0 & 5 \end{bmatrix}, \ |AB| = 5 \begin{vmatrix} 2 & 10 \\ 6 & 33 \end{vmatrix} = 30$$

$$= |A|\,|B|.$$

You can verify Theorem 1 for the following situation.

E7) Show that $|AB| = |A|\,|B|$, where

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & -2 \\ 3 & -3 & 5 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 2 & 0 \\ 5 & -3 & 3 \end{bmatrix}$$

Theorem 1 can be extended to a product of m n×n matrices, $A_1, A_2, \ldots, A_m$. That is,
$\det(A_1 A_2 \ldots A_m) = \det(A_1) \det(A_2) \ldots \det(A_m)$

Now let us look at an example in which Theorem 1 simplifies calculations.

Example 7: For a, b, c ∈ R, calculate

$$\begin{vmatrix} a^2 + 2bc & c^2 + 2ab & b^2 + 2ac \\ b^2 + 2ac & a^2 + 2bc & c^2 + 2ab \\ c^2 + 2ab & b^2 + 2ac & a^2 + 2bc \end{vmatrix}$$

Solution: The solution is very simple. The given matrix is equal to

$$\begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix}^2 . \text{ Therefore,}$$

we get the required determinant to be

$$\left|\begin{bmatrix} a & b & c \\ c & a & b \\ b & c & a \end{bmatrix}^2\right| = \begin{vmatrix} a & b & c \\ c & a & b \\ b & c & a \end{vmatrix}^2 \quad \text{(by Theorem 1)}$$

$$= (a^3+b^3+c^3-3abc)^2,$$

because $\begin{vmatrix} a & b & c \\ c & a & b \\ b & c & a \end{vmatrix} = a \begin{vmatrix} a & b \\ c & a \end{vmatrix} - b \begin{vmatrix} c & b \\ b & a \end{vmatrix} + c \begin{vmatrix} c & a \\ b & c \end{vmatrix}$.

$$= a(a^2-bc) - b(ac-b^2) + c(c^2-ab)$$
$$= a^3 + b^3 + c^3 - 3abc.$$

Now, you know that $AB \neq BA$, in general. But, $\det(AB) = \det(BA)$, since both are equal to the scalar $\det(A)\det(B)$.

On the other hand, $\det(A+B) \neq \det(A) + \det(B)$, in general. The following exercise is an example.

E8) Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. Show that $\det(A+B) \neq \det(A) + \det(B)$.

What we have just said is that det. is not a linear function.

We now give an immediate corollary to Theorem 1.

Corollary 1: If $A \in M_n(F)$ is invertible, then $\det(A^{-1}) = 1/\det(A)$.
Proof: Let $B \in M_n(F)$ such that $AB = I$. Then $\det(AB) = \det(A)\det(B) = \det(I) = 1$. Thus, $\det(A) \neq 0$ and $\det(B) = 1/\det(A)$. In particular, $\det(A^{-1}) = 1/\det(A)$.

Another corollary to Theorem 1 is

Corollary 2: Similar matrices have the same determinant.
Proof: If B is similar to A, then $B = P^{-1}AP$ for some invertible matrix P. Thus, by Theorem 1, $\det(B) = \det(P^{-1}AP)$

$$= \det(P^{-1})\det(A)\det(P) = 1/\det(P).\det(P).\det(A), \text{ by Cor. 1.}$$
$$= \det(A).$$

A matrix B is similar to a matrix A if there exists a non-singular matrix P such that $P^{-1}AP = B$.

We use this corollary to introduce you to the determinant of a linear transformation. At each stage you have seen the very close relationship between linear transformations and matrices. Here too, you will see this closeness.

Definition: Let $T: V \rightarrow V$ be a linear transformation on a finite-dimensional non-zero vector space V. Let $A = [T]_B$ be the matrix of T with respect to a given basis B of V. Then we define the **determinant of T** by $\det(T) = \det(A)$.

This definition is independent of the basis of V that is chosen because, if we choose another basis B' of V we obtain the matrix $A' = [T]_{B'}$, which is similar to A (see Unit 7, Cor. to Theorem 10). Thus, det (A') = det (A).

We have the following example and exercises.

Example 8: Find det(T) where we define $T: \mathbf{R}^3 \to \mathbf{R}^3$ by
$T(x_1,x_2,x_3) = (3x_1+x_3,-2x_1+x_2,-x_1+2x_2+4x_3)$

Solution: Let B = {(1,0,0), (0,1,0), (0,0,1)} be the standard ordered basis of $\mathbf{R}^3$. Now,

T(1,0,0) = (3,-2,-1) = 3(1,0,0) -2(0,1,0) -1(0,0,1)
T(0,1,0) = (0,1,2) = 0(1,0,0) + 1(0,1,0) + 2(0,0,1)
T(0,0,1) = (1,0,4) = 1(1,0,0) + 0(0,1,0) + 4(0,0,1)

$$\therefore, A = [T]_B = \begin{bmatrix} 3 & 0 & 1 \\ -2 & 1 & 0 \\ -1 & 2 & 4 \end{bmatrix}$$

So, by definition,

$$\det(T) = \det(A) = \begin{vmatrix} 3 & 0 & 1 \\ -2 & 1 & 0 \\ -1 & 2 & 4 \end{vmatrix}$$

$$= 3\begin{vmatrix} 1 & 0 \\ 2 & 4 \end{vmatrix} + 1\begin{vmatrix} -2 & 1 \\ -1 & 2 \end{vmatrix} = 12-3 = 9.$$

E E9) Find the determinant of the zero operator and the identity operator from $\mathbf{R}^3 \to \mathbf{R}^3$.

E E10) Consider the differential operator
$D: \mathbf{P}_2 \to \mathbf{P}_2 : D(a_0+a_1x+a_2x^2) = a_1 + 2a_2x$.
What is det(D)?

Let us now see what the adjoint of a square matrix is, and how it will help us in obtaining the inverse of an invertible matrix.

9.4.2 Adjoint of a Matrix

In Section 9.2 we used the notation A_{ij} for the matrix obtained from a square matrix A by deleting its ith row and jth column. Related to this we define the (i,j)th cofactor of A (or the cofactor of a_{ij}) to be $(-1)^{i+j}|A_{ij}|$. It is denoted by C_{ij}. That is $C_{ij} = (-1)^{i+j}|A_{ij}|$.

Consider the following example.

Example 9: Obtain the cofactors C_{12} and C_{23} of the matrix $A = \begin{bmatrix} 0 & 2 & -1 \\ 3 & 4 & 1 \\ 2 & 1 & 6 \end{bmatrix}$.

Solution: $C_{12} = (-1)^{1+2}|A_{12}| = -\begin{vmatrix} 3 & 1 \\ 2 & 6 \end{vmatrix} = -16$

$$C_{23} = (-1)^{2+3}|A_{23}| = -\begin{vmatrix} 0 & 2 \\ 2 & 1 \end{vmatrix} = 4.$$

In the following result we give a relationship between the elements of a matrix and their cofactors.

Theorem 2: Let $A = [a_{ij}]_{n\times n}$. Then,

a) $a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in} = \det(A) = a_{1i}C_{1i} + a_{2i}C_{2i} + \ldots + a_{ni}C_{ni}$.

b) $a_{i1}C_{j1} + a_{i2}C_{j2} + \ldots + a_{in}C_{jn} = 0 = a_{1i}C_{1j} + a_{2i}C_{2j} + \ldots + a_{ni}C_{nj}$ if $i \neq j$.

We will not be proving this theorem here. We only mention that (a) follows immediately from the definition of det (A), since $\det(A) = (-1)^{i+1} a_{i1}|A_{i1}| + \ldots + (-1)^{i+n} a_{in}|A_{in}|$.

E11) Verify (b) of Theorem 2 for the matrix in Example 9 and i=1, j=2 or 3.

Now, we can define the adjoint of a matrix.

Definition: Let $A = [a_{ij}]$ be any n × n matrix. Then the **adjoint** of A is the n × n matrix, denoted by **Adj(A)**, and defined by

$$\text{Adj}(A) = \begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1n} \\ C_{21} & C_{22} & \ldots & C_{2n} \\ \cdot & \cdot & \ldots & \cdot \\ \vdots & \vdots & \ldots & \vdots \\ C_{n1} & C_{n2} & \ldots & C_{nn} \end{bmatrix}^t = \begin{bmatrix} C_{11} & C_{21} & & C_{n1} \\ C_{12} & C_{22} & & C_{n2} \\ \cdot & \cdot & \ldots & \cdot \\ \vdots & \vdots & \ldots & \vdots \\ C_{1n} & C_{2n} & \ldots & C_{nn} \end{bmatrix}.$$

where C_{ij} denotes the (i,j)th cofactor of A.

Thus, Adj(A) is the n × n matrix which is the transpose of the matrix of corresponding cofactors of A.

Let us look at an example.

Example 10: Obtain the adjoint of the matrix $A = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}$

Solution: $C_{11} = (-1)^{1+1} \begin{vmatrix} 1 & 0 \\ 0 & \cos\theta \end{vmatrix} = \cos\theta$

$$C_{12} = (-1)^{1+2} \begin{vmatrix} 0 & 0 \\ \sin\theta & \cos\theta \end{vmatrix} = 0$$

$$C_{13} = \begin{vmatrix} 0 & 1 \\ \sin\theta & 0 \end{vmatrix} = -\sin\theta$$

$C_{21} = 0$, $C_{22} = \cos^2\theta + \sin^2\theta = 1$, $C_{23} = 0$.
$C_{31} = \sin\theta$, $C_{32} = 0$, $C_{33} = \cos\theta$.

$$\therefore \text{Adj}(A) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix}^t = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

Now you can try the following exercise.

E E12) Find Adj(A), where $A = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 0 & 6 \\ 0 & 0 & 5 \end{bmatrix}$.

In Unit 7 you came across one method of finding out if a matrix is invertible. The following theorem uses the adjoint to give another way of finding out if a matrix A is invertible. It also gives us A^{-1}, if A is invertible.

Theorem 3: Let A be an $n \times n$ matrix over F. Then
$A.(\text{Adj}(A)) = (\text{Adj}(A)).A = \det(A)\, I$.

Proof: Recall matrix multiplication from Unit 7. Now

$$A\,(\text{Adj}\,(A)) = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \ldots & C_{n1} \\ C_{12} & C_{22} & \ldots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \ldots & C_{nn} \end{bmatrix}$$

By Theorem 2 we know that $a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in} = \det(A)$, and $a_{i1}C_{j1} + a_{i2}C_{j2} + \ldots + a_{in}C_{jn} = 0$ if $i \neq j$. Therefore,

$$A\,(\text{Adj}\,(A)) = \begin{bmatrix} \det(A) & 0 & \ldots & 0 \\ 0 & \det(A) & & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \det(A) \end{bmatrix}$$

$$= \det(A) \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 \end{bmatrix} = \det(A)\, I.$$

Similarly, (Adj(A)) .A = det (A)I.

An immediate corollary shows us how to calculate the inverse of a matrix, if it exists.

Corollary: Let A be an n × n matrix over F. Then A is invertible if and only if det (A) ≠ 0. If det(A) ≠ 0, then

$A^{-1} = (1/\det(A))\, Adj(A)$,

Proof: If A is invertible, then A^{-1} exists and $A^{-1} A = I$. So, by Theorem 1, $\det(A^{-1}) \det(A) = \det(I) = 1$. ∴, det(A) ≠ 0.

Conversely, if det(A) ≠ 0, then Theorem 3 says that

$$A\left(\frac{1}{|A|} Adj\ (A)\right) = I = \left(\frac{1}{|A|} Adj\ (A)\right) A$$

$$\therefore A^{-1} = \frac{1}{|A|} Adj\ (A).$$

We will use the result in the following example.

Example 11: Let

$$A = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \quad \text{Find } A^{-1}.$$

Solution:

$$\det(A) = (-1)^{2+2}.1. \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} \quad \text{(by expansion along the second row)}$$

$= \cos^2\theta + \sin^2\theta = 1$

Also, from Example 10 we know that

$$Adj(A) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$$

Therefore, $A^{-1} = (1/\det(A))\, Adj(A) = Adj(A)$.

You should also verify that Adj(A) is A^{-1} by calculating A. Adj(A) and Adj(A). A.

You can use Theorem 3 for solving the following exercises.

E13) Can you find A^{-1} for the matrix in E 12?

E14) Find the adjoint and inverse of the matrix A in E7.

E15) If A^{-1} exists, does $[Adj(A)]^{-1}$ exist? If so, what is $[Adj(A)]^{-1}$?

Now we go to the next section, in which we apply our knowledge of determinants to obtain solutions of systems of linear equations.

9.5 SYSTEMS OF LINEAR EQUATIONS

Consider the system of n linear equations in n unknowns, given by

$$a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2$$
$$\vdots$$
$$a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n,$$

which is the same as

$$AX = B, \text{ where } A = [a_{ij}], X = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

In Section 8.4 we discussed the Gaussian elimination method for obtaining a solution of this system. In this section we give a rule due to the mathematician Cramer, for solving a system of linear equations when the number of equations equals the number of variables.

Theorem 4 is called Cramer's Rule.

Theorem 4: Let the matrix equation of a system of linear equations be

$$AX = B, \text{ where } A = [a_{ij}]_{n \times n}, X = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}, B = \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_n \end{bmatrix}$$

Let the columns of A be $C_1, C_2, \ldots, C_n$. If $\det(A) \neq 0$, the given system has a unique solution, namely,

$x_1 = D_1/D, \ldots, x_n = D_n/D$, where

$D_i = \det(C_1, \ldots C_{i-1}, B, C_{i+1}, \ldots, C_n)$

= determinant of the matrix obtained from A by replacing the ith column by B, and

$D = \det(A)$.

Proof: Since $|A| \neq 0$, the corollary to Theorem 3 says that A^{-1} exists.
Now $AX = B \Rightarrow A^{-1}AX = A^{-1}B$
$\Rightarrow IX = (1/D)\,Adj(A)\,B$

$$\Rightarrow X = (1/D)\begin{bmatrix} C_{11} & C_{21} & \dots & C_{n1} \\ C_{12} & C_{22} & \dots & C_{n2} \\ \vdots & \vdots & \dots & \vdots \\ C_{1n} & C_{2n} & \dots & C_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Thus,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = (1/D)\begin{bmatrix} C_{11}b_1 + C_{21}b_2 + \dots + C_{n1}b_n \\ C_{12}b_1 + C_{22}b_2 + \dots + C_{n2}b_n \\ \vdots \\ C_{1n}b_1 + C_{2n}b_2 + \dots + C_{nn}b_n \end{bmatrix}.$$

Now, $D_i = \det(C_1, \dots, C_{i-1}, B, C_{i+1}, \dots, C_n)$. Expanding along the ith column, we get $D_i = C_{1i}b_1 + C_{2i}b_2 + \dots + C_{ni}b_n$.

Thus,

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = 1/D \begin{bmatrix} D_1 \\ D_2 \\ \cdot \\ \cdot \\ D_n \end{bmatrix},$$

which gives us Cramer's Rule, namely,

$x_1 = D_1/D,\ x_2 = D_2/D, \dots, x_n = D_n/D.$

The following example and exercise may help you to practise using Cramer's Rule.

Example 12: Solve the following system using Cramer's Rule:

$$\begin{aligned} 2x + 3y - z &= 2 \\ x + 2y + z &= -1 \\ 2x + y - 6z &= 4 \end{aligned}$$

Solution: The given system is equivalent to $AX = B$, where

$A = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & 1 \\ 2 & 1 & -6 \end{bmatrix}$, $X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, $B = \begin{bmatrix} 2 \\ -1 \\ 4 \end{bmatrix}$. Therefore, applying the rule, we get

$$x = \frac{\begin{vmatrix} 2 & 3 & -1 \\ -1 & 2 & 1 \\ 4 & 1 & -6 \end{vmatrix}}{\begin{vmatrix} 2 & 3 & -1 \\ 1 & 2 & 1 \\ 2 & 1 & -6 \end{vmatrix}},\quad y = \frac{\begin{vmatrix} 2 & 2 & -1 \\ 1 & -1 & 1 \\ 2 & 4 & -6 \end{vmatrix}}{\begin{vmatrix} 2 & 3 & -1 \\ 1 & 2 & 1 \\ 2 & 1 & -6 \end{vmatrix}},\quad z = \frac{\begin{vmatrix} 2 & 3 & 2 \\ 1 & 2 & -1 \\ 2 & 1 & 4 \end{vmatrix}}{\begin{vmatrix} 2 & 3 & -1 \\ 1 & 2 & 1 \\ 2 & 1 & -6 \end{vmatrix}}$$

After calculating, we get
$x = -23,\ y = 14,\ z = -6.$

Substitute these values in the given equations to check that we haven't made a mistake in our calculations.

E16) Solve, by Cramer's Rule, the following system of equations.

$$\begin{aligned} x + 2y + 4z &= 1 \\ 2x + 3y - z &= 3 \\ x - 3z &= 2 \end{aligned}$$

Now let us see what happens if B = 0. Remember, in Unit 8 you saw that **AX = 0** has n − r linearly independent solutions, where r = rank A. The following theorem tells us this condition in terms of det(A).

Theorem 5: The homogeneous system AX = 0 has a non-trivial solution if and only if det(A) = 0.

X is non-trivial if $X \neq 0$.

Proof: First assume that AX = 0 has a non-trivial solution. Suppose, if possible, that $\det(A) \neq 0$. Then Cramer's Rule says that AX = 0 has only the trivial solution X = 0 (because each $D_i = 0$ in Theorem 4). This is a contradiction to our assumption. Therefore, det (A) = 0.

Conversely, if det (A) = 0, then A is not invertible. ∴, the linear mapping $A : V_n(F) \to V_n(F)$: A(X) = AX is not invertible. ∴, this mapping is not one-one. Therefore, $\operatorname{Ker} A \neq \mathbf{0}$, that is $AX = 0$ for some non-zero $X \in V_n(F)$. Thus, **AX = 0** has a non-trivial solution.

You can use Theorem 5 to solve the following exercise.

E17) Does the system

$$\begin{aligned} 2x + 3y + z &= 0 \\ x - y - z &= 0 \\ 4x + 6y + 2z &= 0 \end{aligned}$$

have a non-zero solution?

And now we introduce you to the determinant rank of a matrix, which leads us to another method of obtaining the rank of a matrix.

9.6 THE DETERMINANT RANK

In Units 5 and 8 you were introduced to the rank of a linear transformation and the rank of a matrix, respectively. Then we related the two ranks. In this section we will discuss the determinant rank and show that it is the rank of the concerned matrix. First we give a necessary and sufficient condition for n vectors in $V_n(F)$ to be linearly dependent.

Theorem 6: Let $X_1, X_2, \ldots, X_n \in V_n(F)$. Then $X_1, X_2, \ldots, X_n$ are linearly dependent over the field F if and only if $\det(X_1, X_2, \ldots, X_n) = 0$.

Proof: Let $U = (X_1, X_2, \ldots, X_n)$ be the $n \times n$ matrix whose column vectors are $X_1, X_2, \ldots, X_n$. Then $X_1, X_2, \ldots, X_n$ are linearly dependent over **F** if and only if there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_n \in F$, not all zero, such that $\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n = 0$.

Now,

$$U\begin{bmatrix}\alpha_1\\ \alpha_2\\ \cdot\\ \cdot\\ \alpha_n\end{bmatrix} = (X_1, X_2, \ldots, X_n)\begin{bmatrix}\alpha_1\\ \alpha_2\\ \cdot\\ \cdot\\ \alpha_n\end{bmatrix}$$

$= X_1\alpha_1 + X_2\alpha_2 + \ldots + X_n\alpha_n$
$= \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_n X_n.$

hus, $X_1, X_2, \ldots, X_n$ are linearly dependent over F if and only if $UX = 0$ for some non-ero $X = \begin{bmatrix}\alpha_1\\ \alpha_2\\ \cdot\\ \alpha_n\end{bmatrix} \in V_n(F)$.

ut this happens if and only if $\det(U) = 0$, by Theorem 5. Thus, Theorem 6 is proved.

heorem 6 is equivalent to the statement $X_1, X_2, \ldots, X_n \in V_n(F)$ **are linearly independent if** **nd only if** $\det(X_1, X_2, \ldots, X_n) \neq 0$.

ou can use Theorem 6 for solving the following exercise.

[8) Check if the vectors $\begin{bmatrix}1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\-1\\1\end{bmatrix}, \begin{bmatrix}2\\3\\0\end{bmatrix}$ are linearly independent over **R**.

w, consider the matrix $A = \begin{bmatrix}1 & 2 & 3\\ 0 & 4 & 5\\ 1 & 2 & 3\end{bmatrix}$

ce two rows of A are equal we know that $|A| = 0$. But consider its 2×2 submatrix

A submatrix of A is a matrix that can be obtained from A by deleting some rows and columns.

$A_{13} = \begin{bmatrix} 0 & 4 \\ 1 & 2 \end{bmatrix}$. Its determinant is $-4 \neq 0$. In this case we say that the determinant rank of A is 2.

In general, we have the following definition.

Definition: Let A be an $m \times n$ matrix. If $A \neq 0$, then the **determinant rank of A** is the largest positive integer r such that

i) there exists an $r \times r$ submatrix of A whose determinant is non-zero, and
ii) for $s > r$, the determinant of any $s \times s$ submatrix of A is 0.

Note: The determinant rank r of any $m \times n$ matrix is defined, not only of a square matrix. Also $r \leq \min(m, n)$.

Consider the following example.

Example 13: Obtain the determinant rank of $A = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.

Solution: Since A is a 3×2 matrix, the largest possible value of its determinant rank can be 2. Also, the submatrix $\begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}$ of A has determinant $(-3) \neq 0$.

$\therefore$, the determinant rank of A is 2.

Try the following exercise now.

E E19) Calculate the determinant rank of A, where A =

a) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{bmatrix}$, b) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$.

And now we come to the reason for introducing the determinant rank — it gives us another method for obtaining the rank of a matrix.

Theorem 7: The determinant rank of an $m \times n$ matrix A is equal to the rank of A.

Proof: Let the determinant rank of A be r. Then there exists an $r \times r$ submatrix of A whose determinant is non-zero. By Theorem 6, its column vectors are linearly independent. It follows by the definition of linear independence, that these column vectors, when extended to the column vectors of A, remain linearly independent. Thus, A has at least r linearly independent column vectors. Therefore, by definition of the rank of a matrix,

$r \leq \text{rank}(A) = \rho(A)$ ...... (1)

Also, by definition of $\rho(A)$, we know that the number of linearly independent rows that A has is $\rho(A)$. These rows form a $\rho(A) \times n$ matrix B of rank $\rho(A)$. Thus, B will have $\rho(A)$ linearly independent columns. Retaining these linearly independent columns of B we get a $\rho(A) \times \rho(A)$ submatrix C of B. So, C is a submatrix of A whose determinant will be non-zero, by Theorem 6, since its columns are linearly independent. Thus, by the definition of the determinant rank of A, we get

$\rho(A) \leq r$ .... (2)

(1) and (2) give us us $\rho(A) = r$.

We will use Theorem 7 in the following example.

Example 14: Find the rank of

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 1 & 2 \\ -1 & 2 & 2 \end{bmatrix}.$$

Solution: det (A) = 0. But $\det\left(\begin{bmatrix} 2 & 3 \\ 3 & 1 \end{bmatrix}\right) = -7 \neq 0.$

Thus, by Theorem 7, ρ (A)=2.

Remark: This example shows how Theorem 7 can simplify the calculation of the rank of a matrix in some cases. We don't have to reduce a matrix to echelon form each time. But, in (a) of the following exercise, we see a situation where using this method seems to be as tedious as the row-reduction method.

E E20) Use Theorem 7 to find the rank of A, where A =

a) $\begin{bmatrix} 3 & 1 & 2 & 5 \\ 1 & 2 & -1 & 2 \\ 4 & 3 & 1 & 7 \end{bmatrix}$, b) $\begin{bmatrix} 2 & 3 & 5 & 1 \\ 1 & -1 & 2 & 1 \end{bmatrix}$.

E20 (a) shows how much time can be taken by using this method. On the other hand, E20 (b) shows how little time it takes to obtain $\rho(A)$, using the determinant rank. Thus, the method to be used for obtaining $\rho(A)$ varies from case to case.

We end this unit by briefly mentioning what we have cover in it.

9.7. SUMMARY

In this unit we have covered the following points.

1) The definition of the determinant of a square matrix.
2) The properties P1-P7, of determinants.
3) The statement and use of the fact that $\det(AB) = \det(A)\det(B)$.
4) The definition of the determinant of a linear transformation from U to V, where dim U = dim V.
5) The definition of the adjoint of a square matrix.
(6) The use of adjoints to obtain the inverse of an invertible matrix.
7) The proof and use of Cramer's Rule for solving a system of linear equations.
8) The proof of the fact that the homogeneous system of linear equations AX = 0 has a non-zero solution if and only if $\det(A) = 0$.
9) The definition of the determinant rank, and the proof of the fact that rank of A = determinant rank of A.

9.8 SOLUTIONS/ANSWERS

E1) On expanding by the 2nd row we get

$$|A| = -5|A_{21}| + 4|A_{22}| - |A_{23}|.$$

Now, $|A_{21}| = \begin{vmatrix} 2 & 6 \\ 3 & 2 \end{vmatrix} = 4 - 18 = -14.$

$|A_{22}| = \begin{vmatrix} 1 & 6 \\ 7 & 2 \end{vmatrix} = 2 - 42 = -40.$

$|A_{23}| = \begin{vmatrix} 1 & 2 \\ 7 & 3 \end{vmatrix} = 3 - 14 = -11.$

$\therefore |A| = (-5)(-14) + 4(-40) - (-11) = -79.$

Expanding by the 3rd row, we get

$$|A| = 7|A_{31}| - 3|A_{32}| + 2|A_{33}| = 7\begin{vmatrix} 2 & 6 \\ 4 & 1 \end{vmatrix} - 3\begin{vmatrix} 1 & 6 \\ 5 & 1 \end{vmatrix} + 2\begin{vmatrix} 1 & 2 \\ 5 & 4 \end{vmatrix}$$

$= 7(-22) - 3(-29) + 2(-6) = -79.$

Thus, $|A| = -79$, irrespective of the row that we use to obtain it.

E2) a) $A^t = \begin{bmatrix} 1 & 5 & 7 \\ 2 & 4 & 3 \\ 6 & 1 & 2 \end{bmatrix}$. $\therefore$, on expanding by the first row, we get

$$|A^t| = 1\begin{vmatrix} 4 & 3 \\ 1 & 2 \end{vmatrix} - 5\begin{vmatrix} 2 & 3 \\ 6 & 2 \end{vmatrix} + 7\begin{vmatrix} 2 & 4 \\ 6 & 1 \end{vmatrix} = 5 + 70 + 7(-22) = -79$$

b) $A^t = \begin{bmatrix} -3 & 2 & 1 & 2 \\ -2 & 1 & 0 & 1 \\ 0 & 0 & 1 & -3 \\ 2 & -1 & 2 & 1 \end{bmatrix}$ Since the 3rd row has the maximum number of zeros, we expand along it. Then

$$|A^t| = 1 \begin{vmatrix} -5 & 2 & 2 \\ -2 & 1 & 1 \\ 2 & -1 & 1 \end{vmatrix} - (-3) \begin{vmatrix} -3 & 2 & 1 \\ -2 & 1 & 0 \\ 2 & -1 & 2 \end{vmatrix} = 2+3(2)=8.$$

E3) The magnitude of the required volume is the modulus of

$$\begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

We draw the box in Fig. 2.

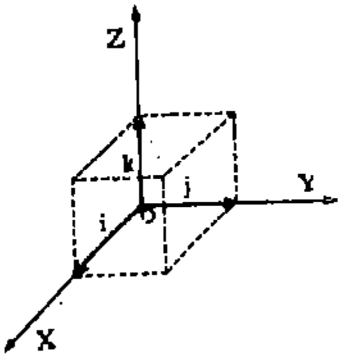


Fig. 2

E4) The first determinant is zero, using the row equivalent of P2. The second determinant is zero, using the row equivalent of P5, since $R_3 = 2R_1$.

E5)
$$\begin{vmatrix} a & 0 & 0 \\ \alpha & b & 0 \\ \beta & \tau & c \end{vmatrix} = a \begin{vmatrix} b & 0 \\ \tau & c \end{vmatrix} = a\,b\,c.$$

$$\begin{vmatrix} a & d & e \\ 0 & b & f \\ 0 & 0 & c \end{vmatrix} = a \begin{bmatrix} b & f \\ 0 & c \end{bmatrix} = a\,b\,c.$$

E6) $\theta(t) = \begin{bmatrix} f(t) & g(t) \\ f'(t) & g'(t) \end{bmatrix} = f(t')\,g'(t) - f'(t)\,g(t).$

$\therefore \theta'(t) = f'(t)\,g'(t) + f(t)\,g''(t) - [f''(t)\,g(t) + f'(t)\,g'(t)],$

since $\frac{d}{dt}(fg) = \frac{df}{dt}g + f\frac{dg}{dt}$

$= f(t)\,g''(t) - f''(t)\,g(t) = \begin{vmatrix} f(t) & g(t) \\ f''(t) & g''(t) \end{vmatrix}$

E7) Note that B is obtained from A by interchanging C_1 and C_3.

$\therefore |B| = -|A|$

Now $|A| = \begin{vmatrix} 2 & -2 \\ -3 & -5 \end{vmatrix} - \begin{vmatrix} 0 & 3 \\ 3 & -3 \end{vmatrix} = 4 + 6 = 10. \quad \therefore |B| = -10.$

Also $|AB| = \begin{vmatrix} -6 & 3 & -2 \\ -14 & 10 & -6 \\ 28 & -21 & 18 \end{vmatrix}$

$= \begin{vmatrix} -6 & 3 & -2 \\ -14 & 10 & -6 \\ 0 & -1 & 6 \end{vmatrix}$ adding $2R_2$ to R_3.

$= \begin{vmatrix} -6 & -2 \\ -14 & -6 \end{vmatrix} + 6 \begin{vmatrix} -6 & 3 \\ -14 & 10 \end{vmatrix}$, expanding along R_3.

$= 8 - 108 = -100 = |A|\,|B|.$

E8) $|A| = 1 = |B|$. $\therefore |A| + |B| = 2.$

But $A + B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ $\therefore |A+B| = 0 \neq |A| + |B|.$

E9) Let B be the standard basis of R^3. The zero operator is

$0 : R^3 \to R^3 : 0(x) = 0 \ \forall\ x \in R^3$. Now, $[0]_B = 0$.

$\therefore \det(0) = 0.$

$I : R^3 \to R^3 : I(x) = x \ \forall\ x \in R^3$, is the identity operator on R^3. Now, $[I]_B = I_3$.

$\therefore \det(I) = \det(I_3) = 1.$

E10) The standard basis for P_2 is $\{1,x,x^2\}$.

Now $D(1) = 0$, $D(x) = 1$, $D(x^2) = 2x$.

$$\therefore [D]_B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\therefore \det(D) = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{vmatrix} = 0$$

E11) $a_{11}C_{21} + a_{12}C_{22} + a_{13}C_{23} = 2(-1)^{2+2} \begin{vmatrix} 0 & -1 \\ 2 & 6 \end{vmatrix} + (-1)(-1)^{2+3} \begin{vmatrix} 0 & 2 \\ 2 & 1 \end{vmatrix} = 0$

Similarly, check that $a_{11}C_{12} + a_{21}C_{22} + a_{31}C_{32} = 0$,

$a_{11}C_{31} + a_{12}C_{32} + a_{13}C_{33} = 0 = a_{11}C_{13} + a_{21}C_{23} + a_{31}C_{33}$.

E12) $C_{11} = 0$, $C_{12} = 0$, $C_{13} = 0$, $C_{21} = -15$, $C_{22} = 10$, $C_{23} = 0$, $C_{31} = 18$, $C_{32} = -12$, $C_{33} = 0$.

$$\therefore \text{Adj}(A) = \begin{bmatrix} 0 & -15 & 18 \\ 0 & 10 & -12 \\ 0 & 0 & 0 \end{bmatrix}.$$

E13) Since $|A| = 0$, A^{-1} does not exist.

E14) From E7 we know that $|A| = 10$.

Now, $C_{11} = 4$, $C_{12} = 6$, $C_{13} = -6$,
$C_{21} = 3$, $C_{22} = 8$, $C_{23} = 3$,
$C_{31} = 2$, $C_{32} = 2$; $C_{33} = 2$.

$$\therefore \text{Adj}(A) = \begin{bmatrix} 4 & 3 & 2 \\ 6 & 8 & 2 \\ -6 & 3 & 2 \end{bmatrix}$$

$$\therefore A^{-1} = \frac{1}{|A|} \text{Adj}(A) = \frac{1}{10} \begin{bmatrix} 4 & 3 & 2 \\ 6 & 8 & 2 \\ -6 & 3 & 2 \end{bmatrix}.$$

Verify that the matrix we have obtained is right, by multiplying it by A.

E15) Since A. Adj (A) = $|A|$ I = Adj (A). A, and $|A| \neq 0$, we find that

$[\text{Adj}(A)]^{-1}$ exists, and is $\frac{1}{|A|}$ A.

E16) This is of the form AX = B, where

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & -1 \\ 1 & 0 & -3 \end{bmatrix}, \; X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \; B = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

$$\therefore D_1 = \begin{vmatrix} 1 & 2 & 4 \\ 3 & 3 & -1 \\ 2 & 0 & -3 \end{vmatrix} = -19$$

$$D_2 = \begin{vmatrix} 1 & 1 & 4 \\ 2 & 3 & -1 \\ 1 & 2 & -3 \end{vmatrix} = 2$$

$$D_3 = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 3 & 3 \\ 1 & 0 & 2 \end{vmatrix} = 1$$

$$D = |A| = -11$$

$$\therefore x = \frac{D_1}{D} = \frac{19}{11}, \; y = \frac{D_2}{D} = \frac{-2}{11}, \; z = \frac{D_3}{D} = \frac{-1}{11}$$

E17) The given system is equivalent to $AX = 0$, where

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & -1 & -1 \\ 4 & 6 & 2 \end{bmatrix}$$

Now, the third row of A is twice the first row of A.

$\therefore$, by P2 and P4 of Section 9.3, $|A| = 0$.

$\therefore$, by Theorem 5, the given system has a non-zero solution.

E18) $\begin{vmatrix} 1 & 0 & 2 \\ 0 & -1 & 3 \\ 1 & 1 & 0 \end{vmatrix} = -3 + 2 = -1 \neq 0$. $\therefore$ the given vectors are linearly independent.

E19) a) Since $|A| \neq 0$, the determinant rank of A is 3.

b) As in Example 13, the determinant rank of A is 2.

E20) a) The determinant rank of $A \leq 3$.
Now the determinant of the 3×3 submatrix $\begin{bmatrix} 3 & 1 & 2 \\ 1 & 2 & -1 \\ 4 & 3 & 1 \end{bmatrix}$ is zero.

Also, the determinant of the 3×3 submatrix $\begin{bmatrix} 3 & 2 & 5 \\ 1 & -1 & 2 \\ 4 & 1 & 7 \end{bmatrix}$ is zero.

In fact, you can check that the determinant of any of the 3×3 submatrices is zero. Now let us look at the 2×2 submatrices of A. Since $\begin{vmatrix} 3 & 1 \\ 1 & 2 \end{vmatrix} = 5 \neq 0$, we find that $\rho(A) = 2$.

b) The determinant rank of $A \leq 2$.

Now $\begin{vmatrix} 2 & 3 \\ 1 & -1 \end{vmatrix} = -5 \neq 0$. $\therefore, \rho(A) = 2$.

UNIT 10 EIGENVALUES AND EIGENVECTORS

Structure

10.1 INTRODUCTION

In Unit 7 you have studied about the matrix of a linear transformation. You have had several opportunities, in earlier units, to observe that the matrix of a linear transformation depends on the choice of the bases of the concerned vector spaces.

Let V be an n-dimensional vector space over F, and let $T : V \to V$ be a linear transformation. In this unit we will consider the problem of finding a suitable basis B, of the vector space V, such that the $n \times n$ matrix $[T]_B$ is a diagonal matrix. This problem can also be seen as: given an $n \times n$ matrix A, find a suitable $n \times n$ non-singular matrix P such that $P^{-1}AP$ is a diagonal matrix (see Unit 7, Cor. to Theorem 10). It is in this context that the study of eigenvalues and eigenvectors plays a central role. This will be seen in Section 10.4.

The eigenvalue problem involves the evaluation of all the eigenvalues and eigenvectors of a linear transformation or a matrix. The solution of this problem has basic applications in almost all branches of the sciences, technology and the social sciences, besides its fundamental role in various branches of pure and applied mathematics. The emergence of computers and the availability of modern computing facilities has further strengthened this study, since they can handle very large systems of equations.

In Section 10.2 we define eigenvalues and eigenvectors. We go on to discuss a method of obtaining them, in Section 10.3. In this section we will also define the characteristic polynomial, of which you will study more in the next unit.

Objectives

After studying this unit, you should be able to

- obtain the characteristic polynomial of a linear transformation or a matrix;
- obtain the eigenvalues, eigenvectors and eigenspaces of a linear transformation or a matrix;
- obtain a basis of a vector space V with respect to which the matrix of a linear transformation $T : V \to V$ is in diagonal form;
- obtain a non-singular matrix P which diagonalises a given diagonalisable matrix A.

10.2 THE ALGEBRAIC EIGENVALUE PROBLEM

Consider the linear mapping $T : R^2 \to R^2 : T(x,y) = (2x, y)$. Then, $T(1,0) = (2,0) = 2(1,0)$. $\therefore T(x,y) = 2(x,y)$ for $(x,y) = (1,0) \neq (0,0)$. In this situation we say that 2 is an eigenvalue of T. But what is an eigenvalue?

λ is the Greek letter 'lambda'.

Definitions: An **eigenvalue** of a linear transformation $T : V \to V$ is a scalar $\lambda \in F$ such that there exists a non-zero vector $x \in V$ satisfying the equation $Tx = \lambda x$.

This non-zero vector $x \in V$ is called an **eigenvector** of T with respect to the eigenvalue λ. (In our example above, (1,0) is an eigenvector of T with respect to the eigenvalue 2.)

Thus, a vector $x \in V$ is an eigenvector of the linear transformation T if

i) x is non-zero, and

ii) $Tx = \lambda x$ for some scalar $\lambda \in F$.

The fundamental algebraic eigenvalue problem deals with the determination of all the eigenvalues of a linear transformation. Let us look at some examples of how we can find eigenvalues.

Example 1: Obtain an eigenvalue and a corresponding eigenvector for the linear operator $T : \mathbf{R}^3 \to \mathbf{R}^3 : T(x,y,z) = (2x, 2y, 2z)$.

Solution: Clearly, $T(x,y,z) = 2(x,y,z) \ \forall \ (x,y,z) \in \mathbf{R}^3$. Thus, 2 is an eigenvalue of T. Any non-zero element of $\mathbf{R}^3$ will be an eigenvector of T corresponding to 2.

Example 2: Obtain an eigenvalue and a corresponding eigenvector of $T : \mathbf{C}^3 \to \mathbf{C}^3 :$ $T(x,y,z) = (ix, -iy, z)$.

Solution: Firstly note that T is a linear operator. Now, if $\lambda \in C$ is an eigenvalue, then $\exists \ (x,y,z) \neq (0,0,0)$ such that $T(x,y,z) = \lambda \ (x, y, z) \Rightarrow (ix, -iy, z) = (\lambda x, \lambda y, \lambda z)$.

$\Rightarrow ix = \lambda x, -iy = \lambda y, z = \lambda z$ .... (1)

These equations are satisfied if $\lambda = i, y = 0, z = 0$.

$\therefore \ \lambda = i$ is an eigenvalue with a corresponding eigenvector being (1,0,0) (or (x,0,0) for any $x \neq 0$).

(1) is also satisfied if $\lambda = -i, x = 0, z = 0$ or if $\lambda = 1, x = 0, y = 0$. Therefore, $-i$ and 1 are also eigenvalues with corresponding eigenvectors (0,y,0) and (0,0,z) respectively, for any $y \neq 0, z \neq 0$.

Do try the following exercise now.

E1) Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be defined by $T(x,y) = (x,0)$. Obtain an eigenvalue and a corresponding eigenvector of T.

Warning: The zero vector can never be an eigenvector. But, $0 \in F$ can be an eigenvalue. For example, 0 is an eigenvalue of the linear operator in E 1, a corresponding eigenvector being (0,1).

Now we define a vector space corresponding to an eigenvalue of $T : V \to V$. Suppose $\lambda \in F$ is an eigenvalue of the linear transformation T. Define the set

$W_\lambda = \{x \in V \mid T(x) = \lambda x\}$

$= \{0\} \cup \{\text{eigenvectors of T corresponding to } \lambda\}$.

So, a vector $v \in W_\lambda$ if and only if $v = 0$ or v is an eigenvector of T corresponding to λ.

Now, $x \in W_\lambda \iff Tx = \lambda Ix$, I being the identity operator.

$\iff (T - \lambda I)x = 0$

$\iff x \in \operatorname{Ker}(T - \lambda I)$

$\therefore$, $W_\lambda = \operatorname{Ker}(T - \lambda I)$, and hence, W_λ is a subspace of V (ref. Unit 5, Theorem 4).

Since λ is an eigenvalue of T, it has an eigenvector, which must be non-zero. Thus, W_λ is non-zero.

Definition: For an eigenvalue λ of T, the non-zero subspace W_λ is called the eigenspace of T associated with the eigenvalue λ.

Let us look at an example.

Example 3: Obtain W_2 for the linear operator given in Example 1.

Solution: $W_2 = \{(x,y,z) \in \mathbf{R}^3 \mid T(x,y,z) = 2(x,y,z)\}$

$= \{(x,y,z) \in \mathbf{R}^3 \mid (2x,2y,2z) = 2(x,y,z)\}$

$\mathbf{R}^3$.

Now, try the following exercise.

E2) For T in Example 2, obtain the complex vector spaces W_i, W_{-i} and W_1.

As with every other concept related to linear transformations, we can define eigenvalues and eigenvectors for matrices also. Let us do so.

Let A be any n×n matrix over the field F.

As we have said in Unit 9 (Theorem 5), the matrix A becomes a linear transformation from $V_n(F)$ to $V_n(F)$, if we define

$A : V_n(F) \to V_n(F) : A(X) = AX.$

Also, you can see that $[A]_{B_0} = A$, where

$$B_0 = \left\{ e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ . \\ . \\ . \\ . \\ 0 \end{bmatrix}, e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ . \\ . \\ . \\ . \\ 0 \end{bmatrix}, \ldots, e_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ . \\ . \\ . \\ 0 \\ 1 \end{bmatrix} \right\}$$

is the standard ordered basis of $V_n(F)$. That is, the matrix of A, regarded as a linear transformation from $V_n(F)$ to $V_n(F)$, with respect to the standard basis B_0, is A itself. This is why we denote the linear transformation A by A itself.

Looking at matrices as linear transformations in the above manner will help you in the understanding of eigenvalues and eigenvectors for matrices.

Definition: A scalar λ is an **eigenvalue of an n×n matrix** A over F if there exists $X \in V_n(F)$, $X \neq 0$, such that $AX = \lambda X$.

If λ is an eigenvalue of A, then all the non-zero vectors in $V_n(F)$ which are solutions of the matrix equation $AX = \lambda X$ are **eigenvectors of the matrix A** corresponding to the eigenvalue λ.

Let us look at a few examples.

Example 4: Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. Obtain an eigenvalue and a corresponding eigenvector of A.

Solution: Now $A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. This shows that 1 is an eigenvalue and $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is an eigenvector corresponding to it.

In fact, $A\begin{bmatrix}0\\1\\0\end{bmatrix} = 2\begin{bmatrix}0\\1\\0\end{bmatrix}$ and $A\begin{bmatrix}0\\0\\1\end{bmatrix} = 3\begin{bmatrix}0\\0\\1\end{bmatrix}$.

Thus, 2 and 3 are eigenvalues of A, with corresponding eigenvectors $\begin{bmatrix}0\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\0\\1\end{bmatrix}$, respectively.

The eigenvalues of diag ($d_1, \ldots\ldots, d_n$) are $d_1, \ldots\ldots, d_n$.

Example 5: Obtain an eigenvalue and a corresponding eigenvector of $A = \begin{bmatrix}0 & -1\\1 & 2\end{bmatrix} \in M_2(R)$.

Solution: Suppose $\lambda \in R$ is an eigenvalue of A. Then

$$\exists\, X = \begin{bmatrix}x\\y\end{bmatrix} \neq \begin{bmatrix}0\\0\end{bmatrix} \text{ such that } AX = \lambda X, \text{ that is, } \begin{bmatrix}-y\\x+2y\end{bmatrix} = \begin{bmatrix}\lambda x\\\lambda y\end{bmatrix}$$

So for what values of λ, x and y are the equations $-y = \lambda x$ and $x+2y = \lambda y$ satisfied? Note that $x \neq 0$ and $y \neq 0$, because if either is zero then the other will have to be zero. Now, solving our equations we get $\lambda = 1$.

Then an eigenvector corresponding to it is $\begin{bmatrix}1\\-1\end{bmatrix}$.

Now you can solve an eigenvalue problem yourself!

E3) Show that 3 is an eigenvalue of $\begin{bmatrix}1 & 2\\0 & 3\end{bmatrix}$. Find 2 corresponding eigenvectors.

Just as we defined an eigenspace associated with a linear transformation we define the **eigenspace** W_λ, corresponding to an eigenvalue λ of an $n \times n$ matrix A, as follows:

$$W_\lambda = \{X \in V_n(F) \mid AX = \lambda X\} = \{X \in V_n(F) \mid (A - \lambda I)X = 0\}$$

For example, the eigenspace W_1, in the situation of Example 4, is

$$\left\{\begin{bmatrix}x\\y\\z\end{bmatrix} \in V_3(R) \,\middle|\, A\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}x\\y\\z\end{bmatrix}\right\} = \left\{\begin{bmatrix}x\\y\\z\end{bmatrix} \in V_3(R) \,\middle|\, \begin{bmatrix}x\\2y\\3z\end{bmatrix} = \begin{bmatrix}x\\y\\z\end{bmatrix}\right\}$$

$$= \left\{\begin{bmatrix}x\\0\\0\end{bmatrix} \,\middle|\, x \in R\right\} \text{ which is the same as } \{(x, 0, 0) \mid x \in R\}.$$

E4) Find W_3 for the matrix in E3.

The algebraic eigenvalue problem for matrices is to determine all the eigenvalues and eigenvectors of a given matrix. In fact, the eigenvalues and eigenvectors of an $n \times n$ matrix A are precisely the eigenvalues and eigenvectors of A regarded as a linear transformation from $V_n(F)$ to $V_n(F)$.

We end this section with the following remark:

A scalar λ is an eigenvalue of the matrix A if and only if $(A - \lambda I) X = 0$ has a non-zero solution, i.e., if and only if $\det (A - \lambda I) = 0$ (by Unit 9, Theorem 5).

Similarly, λ is an eigenvalue of the linear transformation T if and only if $\det (T - \lambda I) = 0$ (ref. Section 9.4).

So far we have been obtaining eigenvalues by observation, or by some calculations that may not give us all the eigenvalues of a given matrix or linear transformation. The remark above suggests where to look for all the eigenvalues In the next section we determine eigenvalues and eigenvectors explicitly.

10.3 OBTAINING EIGENVALUES AND EIGENVECTORS

In the previous section we have seen that a scalar λ is an eigenvalue of a matrix A if and only if $\det (A - \lambda I) = 0$. In this section we shall see how this equation helps us to solve the eigenvalue problem.

10.3.1 Characteristic Polynomial

Once we know that λ is an eigenvalue of a matrix A, the eigenvectors can easily be obtained by finding non-zero solutions of the system of equations given by $AX = \lambda X$.

$$\text{Now, if } A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \text{ and } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

the equation $AX = \lambda X$ becomes

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Writing it out, we get the following system of equations.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= \lambda x_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= \lambda x_2 \\ \cdot \quad \cdot \quad \dots \quad \cdot \quad & \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= \lambda x_n \end{aligned}$$

This is equivalent to the following system.

$$\begin{array}{lcl} (a_{11}-\lambda)x_1 + a_{12}x_2 + \ldots\ldots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + (a_{22}-\lambda)x_2 + \ldots\ldots + a_{2n}x_n & = & 0 \\ \ldots\ldots & & \ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots\ldots + (a_{nn}-\lambda)x_n & = & 0 \end{array}$$

This homogeneous system of linear equations has a non-trivial solution if and only if the determinant of the coefficient matrix is equal to 0 (by Unit 9, Theorem 5). Thus, λ is an eigenvalue of A if and only if

$$\det(A-\lambda I) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \ldots\ldots & a_{1n} \\ a_{21} & a_{22}-\lambda & \ldots\ldots & a_{2n} \\ \cdot & \cdot & \ldots\ldots & \cdot \\ \cdot & \cdot & \ldots\ldots & \cdot \\ a_{n1} & a_{n2} & \ldots\ldots & a_{nn}-\lambda \end{vmatrix} = 0$$

Now, $\det(\lambda I-A) = (-1)^n \det(A-\lambda I)$ (multiplying each row by (-1)). Hence, $\det(\lambda I - A) = 0$ if and only if $\det(A-\lambda I) = 0$.

This leads us to define the concept of the characteristic polynomial.

Definition : Let $A = [a_{ij}]$ be any $n\times n$ matrix. Then the **characteristic polynomial of the matrix A** is defined by

$$f_A(t) = \det(tI - A)$$

$$= \begin{vmatrix} t-a_{11} & -a_{12} & \ldots\ldots & -a_{1n} \\ -a_{21} & t-a_{22} & \ldots\ldots & -a_{2n} \\ & \vdots & \ldots\ldots & \\ -a_{n1} & -a_{n2} & \ldots\ldots & t-a_{nn} \end{vmatrix}$$

$$= t^n + c_1t^{n-1} + c_2t^{n-2} + \ldots\ldots + c_{n-1}t + c_n$$

where the coefficients $c_1, c_2, \ldots\ldots, c_n$ depend on the entries a_{ij} of the matrix A.

The equation $f_A(t) = 0$ is the **characteristic equation** of A.

When no confusion arises, we shall simply write f(t) in place of $f_A(t)$.

Consider the following example.

Example 6: Obtain the characteristic polynomial of the matrix

$$\begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}.$$

Solution: The required polynomial is $\begin{vmatrix} t-1 & -2 \\ 0 & t+1 \end{vmatrix}$

$$= (t-1)(t+1) = t^2 - 1.$$

Now try this exercise.

E5) Obtain the characteristic polynomial of the matrix

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}.$$

Note that λ is an eigenvalue of A iff $\det(\lambda I - A) = f_A(\lambda) = 0$, that is, iff λ **is a root of** the **characteristic polynomial** $f_A(t)$, defined above. Due to this fact, eigenvalues are also called **characteristic roots**, and eigenvectors are called **characteristic vectors.**

The roots of the characteristic polynomial of a matrix A form the set of eigenvalues of A.

For example, the eigenvalues of the matrix in Example 6 are the roots of the polynomial t^2-1, namely, 1 and (– 1).

E E6) Find the eigenvalues of the matrix in E5.

Now, the characteristic polynomial $f_A(t)$ is a polynomial of degree n. Hence, it can have n roots, at the most. Thus, **an n×n matrix has n eigenvalues, at the most.**

For example, the matrix in Example 6 has two eigenvalues, 1 and –1, and the matrix in E5 has 3 eigenvalues.

Now we will prove a theorem that will help us in Section 10.4.

Theorem 1: Similar matrices have the same eigenvalues.

Proof: Let an n×n matrix B be similar to an n×n matrix A.

Then, by definition, $B = P^{-1}AP$, for some invertible matrix P.

Now, the characteristic polynomial of B,

$$\begin{aligned} f_B(t) &= \det(tI-B) \\ &= \det(tI-P^{-1}AP) \\ &= \det(P^{-1}(tI-A)P), \text{ since } P^{-1}tIP = tP^{-1}P = tI. \\ &= \det(P^{-1})\det(tI-A)\det(P) \text{ (see Sec. 9.4)} \\ &= \det(tI-A)\det(P^{-1})\det(P) \\ &= f_A(t)\det(P^{-1}P) \\ &= f_A(t), \text{ since } \det(P^{-1}P) = \det(I) = 1. \end{aligned}$$

Thus, the roots of $f_B(t)$ and $f_A(t)$ coincide. Therefore, the eigenvalues of A and B are the same.

Let us consider some more examples so that the concepts mentioned in this section become absolutely clear to you.

Example 7 : Find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix}$$

Solution: In solving E6 you found that the eigenvalues of A are $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -2$. Now we obtain the eigenvectors of A.

The eigenvectors of A with respect to the eigenvalue $\lambda_1 = 1$ are the non-trivial solutions of

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

which gives the equations

$$\left.\begin{aligned} 2x_3 &= x_1 \\ x_1 + x_3 &= x_2 \\ x_2 - 2x_3 &= x_3 \end{aligned}\right\} \Longrightarrow \begin{aligned} x_1 &= 2x_3 \\ x_2 &= x_1 + x_3 = 3x_3 \\ x_3 &= x_3 \end{aligned}$$

Thus,

$\begin{bmatrix} 2x_3 \\ 3x_3 \\ x_3 \end{bmatrix}$, that is, $x_3\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ gives all the eigenvectors associated with the eigenvalue $\lambda_1 = 1$, as x_3 takes all non-zero real values.

The eigenvectors of A with respect to $\lambda_2 = -1$ are the non-trivial solutions of

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (-1)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which gives

$$\left.\begin{aligned} 2x_3 &= -x_1 \\ x_1 + x_3 &= -x_2 \\ x_2 - 2x_3 &= -x_3 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= -2x_3 \\ x_2 &= -x_1 - x_3 = 2x_3 - x_3 = x_3 \\ x_3 &= x_3 \end{aligned}$$

Thus, the eigenvectors associated with (–1) are

$$\begin{bmatrix} -2x_3 \\ x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \quad \forall\ x_3 \neq 0, x_3 \in \mathbf{R}.$$

The eigenvectors of A with respect to $\lambda_3 = -2$ are given by

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (-2)\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which gives

$$\left.\begin{aligned} 2x_3 &= -2x_1 \\ x_1 + x_3 &= -2x_2 \\ x_2 - 2x_3 &= -2x_3 \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= -x_3 \\ x_2 &= 0 \\ x_3 &= x_3 \end{aligned}$$

Thus, the eigenvectors corresponding to –2 are

$$\begin{bmatrix} -x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \ x_3 \neq 0, x_3 \in \mathbf{R}.$$

Thus, in this example, the eigenspaces W_1, W_{-1} and W_{-2} are 1- dimensional spaces, generated over **R** by

$$\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \text{ and } \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \text{respectively.}$$

Example 8: Let A be the 4×4 real matrix

$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$. Obtain its eigenvalues and eigenvectors.

Solution: The characteristic polynomial of $A = f_A(t) = \det(tI-A)$

$$= \begin{vmatrix} t-1 & -1 & 0 & 0 \\ 1 & t+1 & 0 & 0 \\ 2 & 2 & t-2 & -1 \\ -1 & -1 & 1 & t \end{vmatrix} = t^2(t-1)^2$$

Thus, the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 1$.

The eigenvectors corresponding to $\lambda_1 = 0$ are given by

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix},$$

which gives $x_1 + x_2 = 0$

$-x_1 - x_2 = 0$

$-2x_1 - 2x_2 + 2x_3 + x_4 = 0$

$x_1 + x_2 - x_3 = 0$

The first and last equations give $x_3 = 0$. Then, the third equation gives $x_4 = 0$. The first equation gives $x_1 = -x_2$.

Thus, the eigenvectors are

$$\begin{bmatrix} -x_2 \\ x_2 \\ 0 \\ 0 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_2 \neq 0, x_2 \in R.$$

The eigenvectors corresponding to $\lambda_2 = 1$ are given by

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -2 & -2 & 2 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix},$$

which gives $x_1 + x_2 = x_1$

$-x_1 - x_2 = x_2$

$-2x_1 - 2x_2 + 2x_3 + x_4 = x_3$

$x_1 + x_2 - x_3 = x_4$

The first two equations give $x_2 = 0$ and $x_1 = 0$. Then the last equation gives $x_4 = -x_3$. Thus, the eigenvectors are

$$\begin{bmatrix} 0 \\ 0 \\ x_3 \\ -x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, x_3 \neq 0, x_3 \in R.$$

Example 9 : Obtain the eigenvalues and eigenvectors of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Solution: The characteristic polynomial of $A = f_A(t) = \det(tI-A)$

$$= \begin{vmatrix} t & -1 & 0 \\ -1 & t & 0 \\ 0 & 0 & t-1 \end{vmatrix} = (t+1)(t-1)^2$$

Therefore, the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = 1$.

The eigenvectors corresponding to $\lambda_1 = -1$ are given by

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (-1) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

which is equivalent to

$x_2 = -x_1$

$x_1 = -x_2$

$x_3 = -x_3$

The last equation gives $x_3 = 0$. Thus, the eigenvectors are

$$\begin{bmatrix} x_1 \\ -x_1 \\ 0 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, x_1 \neq 0, x_1 \in \mathbf{R}.$$

The eigenvectors corresponding to $\lambda_2 = 1$ are given by

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

which gives $x_2 = x_1$
$x_1 = x_2$
$x_3 = x_3$

Thus, the eigenvectors are

$$\begin{bmatrix} x_1 \\ x_1 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

where x_1, x_3 are real numbers, not simultaneously 0.
Note that, corresponding to $\lambda_2 = 1$, there exist two linearly independent eigenvectors $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, which form a basis of the eigenspace W_1.

Thus, W_{-1} is 1-dimensional, while $\dim_{\mathbf{R}} W_1 = 2$.

Try the following exercises now.

E E7) Find the eigenvalues and bases for the eigenspaces of the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix}$$

E8) Find the eigenvalues and eigenvectors of the diagonal matrix

$$D = \begin{bmatrix} a_1 & 0 & 0 & . & . & 0 \\ 0 & a_2 & 0 & . & . & 0 \\ 0 & 0 & a_3 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & a_n \end{bmatrix}, \text{ where } a_i \neq a_j \text{ for } i \neq j.$$

We now turn to the eigenvalues and eigenvectors of linear transformations.

10.3.2 Eigenvalues of Linear Transformations

As in Section 10.2, let $T : V \to V$ be a linear transformation on a finite-dimensional vector space V over the field F. We have seen that
$\lambda \in F$ is an eigenvalue of T

$\Longleftrightarrow \det(T - \lambda I) = 0$
$\Longleftrightarrow \det(\lambda I - T) = 0$
$\Longleftrightarrow \det(\lambda I - A) = 0$, where $A = [T]_B$ is the matrix of T with respect to a basis B of V.
Note that $[\lambda I - T]_B = \lambda I - [T]_B$.

This shows that λ is an eigenvalue of T if and only if λ is an eigenvalue of the matrix $A = [T]_B$, where B is a basis of V. We define the **characteristic polynomial of the linear transformation T** to be the same as the characteristic polynomial of the matrix $A = [T]_B$, where B is a basis of V.

This definition does not depend on the basis B chosen, since similar matrices have the same characteristic polynomial (Theorem 1), and the matrices of the same linear transformation T with respect to two different ordered bases of V are similar.

Just as for matrices, the eigenvalues of T are precisely the roots of the characteristic polynomial of T.

Example 10: Let $T : \mathbf{R}^2 \to \mathbf{R}^2$ be the linear transformation which maps $e_1 = (1,0)$ to $e_2 = (0,1)$ and e_2 to $-e_1$. Obtain the eigenvalues of T.

Solution: Let $A = [T]_B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, where $B = \{e_1, e_2\}$.

The characteristic polynomial of T = the characteristic polynomial of A

$$= \begin{vmatrix} t & 1 \\ -1 & t \end{vmatrix} = t^2 + 1,$$ which has no real roots.

Hence, the linear transformation T has no real eigenvalues. But, it has two complex eigenvalues i and $-i$.

Try the following exercises now.

E9) Obtain the eigenvalues and eigenvectors of the differential operator $D : \mathbf{P}_2 \to \mathbf{P}_2$: $D(a_0 + a_1x + a_2x^2) = a_1 + 2a_2x$, for $a_0, a_1, a_2 \in \mathbf{R}$.

E10) Show that the eigenvalues of a square matrix A coincide with those of A^t.

E11) Let A be an invertible matrix. If λ is an eigenvalue of A, show that $\lambda \neq 0$ and that λ^{-1} is an eigenvalue of A^{-1}.

Now that we have discussed a method of obtaining the eigenvalues and eigenvectors of a matrix, let us see how they help in transforming any square matrix into a diagonal matrix.

10.4 DIAGONALISATION

In this section we start with proving a theorem that discusses the linear independence of eigenvectors corresponding to different eigenvalues.

Theorem 2 : Let $T : V \to V$ be a linear transformation on a finite-dimensional vector space V over the field F. Let $\lambda_1, \lambda_2, .., \lambda_m$ be the distinct eigenvalues of T and $v_1, v_2, ...,$ v_m be eigenvectors of T corresponding to $\lambda_1, \lambda_2, ..., \lambda_m$, respectively. Then $v_1, v_2, ...,$ v_m are linearly independent over F.

Proof : We know that

$Tv_i = \lambda_i v_i, \lambda_i \in F, 0 \neq v_i \in V$ for $i = 1, 2, ..., m$, and $\lambda_i \neq \lambda_j$ for $i \neq j$.

Suppose, if possible, that $\{v_1, v_2, ..., v_m\}$ is a linearly dependent set. Now, the single non-zero vector v_1 is linearly independent. We choose $r(\leq m)$ such that $\{v_1, v_2, ..., v_{r-1}\}$ is linearly independent and $\{v_1, v_2, ..., v_{r-1}, v_r\}$ is linearly dependent. Then

$v_r = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_{r-1} v_{r-1}$ ....(1)

for some $\alpha_1, \alpha_2, ..., \alpha_{r-1}$ in F.

Applying T, we get

$Tv_r = \alpha_1 T v_1 + \alpha_2 T v_2 + ... + \alpha_{r-1} T v_{r-1}$. This gives

$\lambda_r v_r = \alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + ... + \alpha_{r-1} \lambda_{r-1} v_{r-1}$ .... (2)

Now, we multiply (1) by λ_r and subtract it from (2), to get

$0 = \alpha_1(\lambda_1 - \lambda_r) v_1 + \alpha_2(\lambda_2 - \lambda_r) v_2 + ... + \alpha_{r-1} (\lambda_{r-1} - \lambda_r) v_{r-1}$

Since the set $\{v_1, v_2, ..., v_{r-1}\}$ is linearly independent, each of the coefficients in the above equation must be 0. Thus, we have $\alpha_i(\lambda_i - \lambda_r) = 0$ for $i = 1, 2, ... r-1$.

But $\lambda_i \neq \lambda_r$ for $i = 1, 2, ... r-1$. Hence $(\lambda_i - \lambda_r) \neq 0$ for $i = 1, 2 ......, r-1$, and we must have $\alpha_i = 0$ for $i = 1, 2, ..., r-1$. However, this is not possible since (1) would imply that $v_r = 0$, and, being an eigenvector, v_r can never be 0. Thus, we reach a contradiction.

Hence, the assumption we started with must be wrong. Thus, $\{v_1, v_2, ..., v_m\}$ must be linearly independent, and the theorem is proved.

We will use Theorem 2 to choose a basis for a vector space V so that the matrix $[T]_B$ is a diagonal matrix.

Definition : A linear transformation $T : V \to V$ on a finite-dimensional vector space V is said to be **diagonalisable** if there exists a basis $B = \{v_1, v_2, ..., v_n\}$ of V such that the matrix of T with respect to the basis B is diagonal. That is,

$$[T]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & .... & 0 \\ 0 & \lambda_2 & 0 & ... & 0 \\ 0 & 0 & \lambda_3 & ... & . \\ . & . & . & ... & . \\ 0 & 0 & 0 & ... & \lambda_n \end{bmatrix},$$

where $\lambda_1, \lambda_2, ..., \lambda_n$ are scalars which need not be distinct.

The next theorem tells us under what conditions a linear transformation is diagonalisable.

Theorem 3 : A linear transformation T, on a finite-dimensional vector space V, is diagonalisable if and only if there exists a basis of V consisting of eigenvectors of T.

Proof: Suppose that T is diagonalisable. By definition, there exists a basis $B = \{v_1, v_2, ..., v_n\}$ of V, such that

$$[T]_B = \begin{bmatrix} \lambda_1 & 0 & 0 & .... & 0 \\ 0 & \lambda_2 & 0 & .... & 0 \\ 0 & 0 & \lambda_3 & .... & . \\ .. & .. & .. & .... & . \\ 0 & 0 & 0 & .... & \lambda_n \end{bmatrix},$$

By definition of $[T]_B$, we must have

$Tv_1 = \lambda_1 v_1, Tv_2 = \lambda_2 v_2, ..., Tv_n = \lambda_n v_n$.

Since basis vectors are always non-zero, $v_1, v_2, ..., v_n$ are non-zero. Thus, we find that $v_1, v_2, ..., v_n$ are eigenvectors of T.

Conversely, let $B = \{v_1, v_2, ..., v_n\}$ be a basis of V consisting of eigenvectors of T. Then, there exist scalars $\alpha_1, \alpha_2, ..., \alpha_n$, not necessarily distinct, such that $Tv_1 = \alpha_1 v_1$, $Tv_2 =$ $\alpha_2 v_2, ..., Tv_n = \alpha_n v_n$.

But then we have

$$[T]_B = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ 0 & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \alpha_n \end{bmatrix}$$, which means that T is diagonalisable.

The next theorem combines Theorems 2 and 3.

Theorem 4: Let $T: V \to V$ be a linear transformation, where V is an n-dimensional vector space. Assume that T has n distinct eigenvalues. Then T is diagonalisable.

Proof: Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the n distinct eigenvalues of T. Then there exist eigenvectors $v_1, v_2, \dots, v_n$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. By Theorem 2, the set $v_1, v_2, \dots, v_n$, is linearly independent and has n vectors, where n = dim V. Thus, from Unit 5 (corollary to Theorem 5), $B = \{v_1, v_2, \dots, v_n\}$ is a basis of V consisting of eigenvectors of T. Thus, by Theorem 3, T is diagonalisable.

Just as we have reached the conclusions of Theorem 4 for linear transformations, we define diagonalisability of a matrix, and reach a similar conclusion for matrices.

Definition: An $n \times n$ matrix A is said to be diagonalisable if A is similar to a diagonal matrix, that is, $P^{-1}AP$ is diagonal for some non-singular $n \times n$ matrix P.

Note that the matrix A is diagonalisable if and only if the matrix A, regarded as a linear transformation $A: V_n(F) \to V_n(F) : A(X) = AX$, is diagonalisable.

Thus, Theorems 2, 3 and 4 are true for the matrix A regarded as a linear transformation from $V_n(F)$ to $V_n(F)$. Therefore, given an $n \times n$ matrix A, we know that it is diagonalisable if it has n distinct eigenvalues.

We now give a practical method of diagonalising a matrix.

Theorem 5: Let A be an $n \times n$ matrix having n distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Let $X_1, X_2, \dots, X_n \in V_n(F)$ be eigenvectors of A corresponding to $\lambda_1, \lambda_2, \dots, \lambda_n$, respectively. Let $P = (X_1, X_2, \dots, X_n)$ be the $n \times n$ matrix having $X_1, X_2, \dots, X_n$ as its column vectors. Then

$P^{-1}AP = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

Proof: By actual multiplication, you can see that

$$\begin{aligned} AP &= A(X_1, X_2, \dots, X_n) \\ &= (AX_1, AX_2, \dots, AX_n) \\ &= (\lambda_1 X_1, \lambda_2 X_2, \dots, \lambda_n X_n) \\ &= (X_1, X_2, \dots, X_n) \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ . & . & \dots & . \\ . & . & \dots & . \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \\ &= P\,\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n). \end{aligned}$$

Now, by Theorem 2, the column vectors of P are linearly independent. This means that P is invertible (Unit 9, Theorem 6). Therefore, we can pre-multiply both sides of the matrix equation $AP = P\,\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ by P^{-1} to get $P^{-1}AP = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.

Let us see how this theorem works in practice.

Example 11: Diagonalise the matrix

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & -6 \\ 2 & -2 & 3 \end{bmatrix}.$$

Solution: The characteristic polynomial of $A = f(t) =$

$$\begin{vmatrix} t-1 & -2 & 0 \\ -2 & t-1 & 6 \\ -2 & 2 & t-3 \end{vmatrix} = (t-5)(t-3)(t+3).$$

Thus, the eigenvalues of A are $\lambda_1 = 5, \lambda_2 = 3, \lambda_3 = -3$ Since they are all distinct, A is diagonalisable (by Theorem 4). You can find the eigenvectors by the method already explained to you. Right now you can directly verify that

$$A \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, A \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ and } A \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} = -3 \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Thus, $\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$

are eigenvectors corresponding to the distinct eigenvalues 5, 3 and –3, respectively. By Theorem 5, the matrix which diagonalises A is given by

$P = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 1 & 2 \\ -1 & 0 & 1 \end{bmatrix}$. Check, by actual multiplication, that

$P^{-1}AP = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix}$, which is in diagonal form.

The following exercise will give you some practice in diagonalising matrices.

E E12) Are the matrices in Examples 7, 8 and 9 diagonalisable? If so, diagonalise them.

We end this unit by summarising what has been done in it.

10.5 SUMMARY

As in previous units, in this unit also we have treated linear transformations along with the analogous matrix version. We have covered the following points here.

1) The definition of eigenvalues, eigenvectors and eigenspaces of linear transformations and matrices.
2) The definition of the characteristic polynomial and characteristic equation of a linear transformation (or matrix).
3) A scalar λ is an eigenvalue of a linear transformation T(or matrix A) if and only if it is a root of the characteristic polynomial of T (or A).
4) A method of obtaining all the eigenvalues and eigenvectors of a linear transformation (or matrix).
5) Eigenvectors of a linear transformation (or matrix) corresponding to distinct eigenvalues are linearly independent.
6) A linear transformation $T : V \rightarrow V$ is diagonalisable if and only if V has a basis consisting of eigenvectors of T.
7) A linear transformation (or matrix) is diagonalisable if its eigenvalues are distinct.

10.6 SOLUTIONS/ANSWERS

E1) Suppose $\lambda \in \mathbf{R}$ is an eigenvalue. Then $\exists$ $(x,y) \neq (0,0)$ such that $T(x,y) = \lambda (x, y)$ $\Rightarrow (x,0) = (\lambda x, \lambda y) \Rightarrow \lambda x = x$, $\lambda y = 0$. These equations are satisfied if $\lambda = 1$, $y = 0$. $\therefore$, 1 is an eigenvalue. A corresponding eigenvector is (1,0). Note that there are infinitely many eigenvectors corresponding to 1, namely, $(x, 0) \ \forall \ 0 \neq x \in \mathbf{R}$.

E2) $W_i = \{(x, y, z) \in C^3 \mid T(x, y, z) = i(x, y, z)\}$

$= \{(x, y, z) \in C^3 \mid (ix, -iy, z) = (ix, iy, iz)\}$

$= \{(x, 0, 0) \mid x \in C\}$.

Similarly, you can show that $W_{-i} = \{(0, x, 0) \mid x \in \mathbf{C}\}$ and $W_1 = \{(0, 0, x) \mid x \in \mathbf{C}\}$.

E 3) If 3 is an eigenvalve, then $\exists \begin{bmatrix} x \\ y \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ such that

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 3 \begin{bmatrix} x \\ y \end{bmatrix} \Rightarrow x + 2y = 3x \text{ and } 3y = 3y.$$

These equations are satisfied by $x = 1, y = 1$ and $x = 2, y = 2$.

$\therefore$ 3 is an eigenvalue, and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ are eigenvectors corresponding to 3.

E4) $$W_3 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in V_2(R) \;\middle|\; \begin{bmatrix} x+2y \\ 3y \end{bmatrix} = \begin{bmatrix} 3x \\ 3y \end{bmatrix} \right\}$$

$$= \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in V_2(R) \;\middle|\; x = y \right\} = \left\{ \begin{bmatrix} x \\ x \end{bmatrix} \;\middle|\; x \in R \right\}$$

This is the 1-dimensional real subspace of $V_2(R)$ whose basis is $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$

W5) It is $$\begin{vmatrix} t & 0 & -2 \\ -1 & t & -1 \\ 0 & -1 & t+2 \end{vmatrix} = t \begin{vmatrix} t & -1 \\ -1 & t+2 \end{vmatrix} + \begin{vmatrix} 0 & -2 \\ -1 & t+2 \end{vmatrix}$$

$$= \{t^2(t+2)-t\} - 2 = t^3 + 2t^2 - t - 2.$$

E6) The eigenvalues are the roots of the polynomial $t^3+2t^2-t-2 = (t-1)(t+1)(t+2)$
$\therefore$, they are 1, −1, −2.

E7) $f_A(t) = \begin{vmatrix} t-2 & -1 & 0 \\ 0 & t-1 & 1 \\ 0 & -2 & t-4 \end{vmatrix} = (t-2)^2(t-3)$

$\therefore$, the eigenvalues are $\lambda_1 = 2$, $\lambda_2 = 3$.

The eigenvectors corresponding to λ_1 are given by

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 2 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

This leads us to the equations

$$\left.\begin{aligned} 2x + y &= 2x \\ y - z &= 2y \\ 2y + 4z &= 2z \end{aligned}\right\} \Longrightarrow \begin{aligned} x &= x \\ y &= 0 \\ z &= 0 \end{aligned}$$

$\therefore W_2 = \left\{ \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} \middle| x \in \mathbf{R} \right\}$. $\therefore$, a basis for W_2 is $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$.

The eigenvectors corresponding to λ_2 are given by

$$\begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 3 \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

. This gives us the equations

$$\left.\begin{aligned} 2x + y &= 3x \\ y - z &= 3y \\ 2y + 4z &= 3z \end{aligned}\right\} \Longrightarrow \begin{aligned} x &= x \\ y &= x \\ z &= -2x \end{aligned}$$

$\therefore W_3 = \left\{ \begin{bmatrix} x \\ x \\ -2x \end{bmatrix} \middle| x \in \mathbf{R} \right\}$. $\therefore$, a basis for W_3 is $\left\{ \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$.

E8) $f_D(t) = \begin{vmatrix} t-a_1 & 0 & \dots & 0 \\ 0 & t-a_2 & \dots & 0 \\ . & . & \dots & . \\ . & . & \dots & . \\ 0 & 0 & \dots & t-a_n \end{vmatrix} = (t-a_1)(t-a_2)\dots(t-a_n)$

$\therefore$, its eigenvalues are $a_1, a_2, \dots, a_n$.
The eigenvectors corresponding to a_i are given by

$$\begin{bmatrix} a_1 & 0 & \dots & 0 \\ 0 & a_2 & \dots & 0 \\ . & . & \dots & : \\ 0 & 0 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_n \end{bmatrix} = a_1 \begin{bmatrix} x_1 \\ x_2 \\ : \\ x_n \end{bmatrix}$$

This gives us the equations

$$\left\{\begin{aligned} a_1x_1 &= a_1x_1 \\ a_2x_2 &= a_1x_2 \\ &\vdots \\ a_nx_n &= a_1x_n \end{aligned}\right\} \Rightarrow \begin{aligned} x_1 &= x_1 \\ x_2 &= 0 \\ &\vdots \\ x_n &= 0 \end{aligned}$$

(since $a_i \neq a_j$ for $i \neq j$).

$\therefore$, the eigenvectors corresponding to a_1 are $\begin{bmatrix} x_1 \\ 0 \\ : \\ 0 \end{bmatrix}$, $x_1 \neq 0$, $x_1 \in \mathbf{R}$.

Similarly, the eigenvectors corresponding to a_i are $\begin{bmatrix} 0 \\ \vdots \\ 0 \\ x_i \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $x_i \neq 0, x_i \in \mathbb{R}$.

E9) $B = \{1, x, x^2\}$ is a basis of P_2.

Then $[D]_B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$.

$\therefore$, the characteristic polynomial of D is $\begin{vmatrix} t & -1 & 0 \\ 0 & t & -2 \\ 0 & 0 & t \end{vmatrix} = t^3$

$\therefore$, the only eigenvalue of D is $\lambda = 0$.
The eigenvectors corresponding to $\lambda = 0$ are $a_0 + a_1x + a_2x^2$, where $D(a_0 + a_1x + a_2x^2) = 0$, that is, $a_1 + 2a_2x = 0$.
This gives $a_1 = 0$, $a_2 = 0$. $\therefore$, the set of eigenvectors corresponding to $\lambda = 0$ are $\{a_0 | a_0 \in \mathbf{R}, a_0 \neq 0\} = \mathbf{R} \setminus \{0\}$.

E10) $|tI - A| = |(tI - A)^t|$, since $|A^t| = |A|$.
$= |tI - A^t|$, since $I^t = I$ and $(B - C)^t = B^t - C^t$.
$\therefore$, the eigenvalues of A are the same as those of A^t.

E11) Let X be an eigenvector corresponding to λ. Then $X \neq 0$ and $AX = \lambda X$.
$\therefore A^{-1}(AX) = A^{-1}(\lambda X)$.
$\Rightarrow (A^{-1}A)X = \lambda(A^{-1}X)$
$\Rightarrow X = \lambda(A^{-1}X)$
$\Rightarrow \lambda \neq 0$, since $X \neq 0$.
Also, $X = \lambda(A^{-1}X) \Rightarrow \lambda^{-1}X = A^{-1}X \Rightarrow \lambda^{-1}$ is an eigenvalue of A^{-1}.

E12) Since the matrix in Example 7 has distinct eigenvalues 1, –1 and –2, it is diagonalisable. Eigenvectors corresponding to

these eigenvalues are $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$, respectively.

$\therefore$, if $P = \begin{bmatrix} 2 & -2 & -1 \\ 3 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$, then $P^{-1} \begin{bmatrix} 0 & 0 & 2 \\ 1 & 0 & 1 \\ 0 & 1 & -2 \end{bmatrix} P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}$.

The matrix in Example 8 is not diagonalisable. This is because it only has two distinct eigenvalues and, corresponding to each, it has only one linearly independent eigenvector. $\therefore$, we cannot find a basis of $V_3(F)$ consisting of eigenvectors. And now apply Theorem 3.

The matrix in Example 9 is diagonalisable though it only has two distinct eigenvalues. This is because corresponding to $\lambda_1 = -1$ there is one linearly independent eigenvector, but corresponding to $\lambda_2 = 1$ there exist two linearly independent eigenvectors. Therefore, we can form a basis of $V_3(\mathbb{R})$ consisting of the eigenvectors

$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

The matrix $P = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is invertible, and

$P^{-1} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} P = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

UNIT 11 CHARACTERISTIC AND MINIMAL POLYNOMIAL

Structure

11.1 INTRODUCTION

This unit is basically a continuation of the previous unit, but the emphasis is on a different aspect of the problem discussed in the previous unit.

Let $T:V \to V$ be a linear transformation on an n-dimensional vector space V over the field F. The two most important polynomials that are associated with T are the characteristic polynomial of T and the minimal polynomial of T. We defined the former in Unit 10 and the latter in Unit 6.

In this unit we first show that every square matrix (or linear transformation $T:V \to V$) satisfies its characteristic equation, and use this to compute the inverse of the concerned matrix (or linear transformation), if it exists.

Then we define the minimal polynomial of a square matrix, and discuss the relationship between the characteristic and minimal polynomials. This leads us to a simple way of obtaining the minimal polynomial of a matrix (or linear transformation).

We advise you to study Units 6, 9 and 10 before starting this unit.

Objectives

After studying this unit, you should be able to

- state and prove the Cayley-Hamilton theorem;
- find the inverse of an invertible matrix using this theorem;
- prove that a scalar λ is an eigenvalue if and only if it is a root of the minimal polynomial;
- obtain the minimal polynomial of a matrix (or linear transformation) if the characteristic polynomial is known.

11.2 CAYLEY-HAMILTON THEOREM

In this section we present the Cayley-Hamilton theorem, which is related to the characteristic equation of a matrix. It is named after the British mathematicians Arthur Cayley (1821-1895) and William Hamilton (1805-1865), who were responsible for a lot of work done in the theory of determinants.

Let us consider the 3×3 matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 1 \\ 0 & 3 & 2 \end{bmatrix}$

Then $tI - A = \begin{bmatrix} t & -1 & -2 \\ 1 & t-2 & -1 \\ 0 & -3 & t-2 \end{bmatrix}$.

Let Δ_{ij} be the (i,j)th cofactor of $(tI - A)$.

Then $\Delta_{11} = (t-2)^2 - 3 = t^2 - 4t + 1$, $\Delta_{12} = t-2$, $\Delta_{13} = -3$, $\Delta_{21} = t+4$, $\Delta_{22} = t^2 - 2t$, $\Delta_{23} = 3t$, $\Delta_{31} = 2t-3$, $\Delta_{32} = t-2$, $\Delta_{33} = t^2-2t+1$.

$$\therefore \text{Adj}(tI - A) = \begin{bmatrix} t^2 - 4t + 1 & t + 4 & 2t - 3 \\ t - 2 & t^2 - 2t & t - 2 \\ -3 & 3t & t^2 - 2t + 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} t^2 + \begin{bmatrix} -4 & 1 & 2 \\ 1 & -2 & 1 \\ 0 & 3 & -2 \end{bmatrix} t + \begin{bmatrix} 1 & 4 & -3 \\ -2 & 0 & -2 \\ -3 & 0 & 1 \end{bmatrix}$$

This is a polynomial in t of degree 2, with matrix coefficients.

Similarly, if we consider the $n \times n$ matrix $A = [a_{ij}]$, then $\text{Adj}(tI - A)$ is a polynomial of degree $\leq n-1$, with matrix coefficients. Let

$$\text{Adj}(tI - A) = B_1 t^{n-1} + B_2 t^{n-2} + \ldots + B_{n-1} t + B_n \qquad \ldots (1)$$

where $B_1, \ldots, B_n$ are $n \times n$ matrices over F.

Now, the characteristic polynomial of A is given by

$f(t) = f_A(t) = \det(tI-A) = |tI-A|$

$$= \begin{vmatrix} t-a_{11} & -a_{12} & \cdot & \cdot & \cdot & -a_{1n} \\ -a_{21} & t-a_{22} & \cdot & \cdot & \cdot & -a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n1} & -a_{n2} & \cdot & \cdot & \cdot & t-a_{nn} \end{vmatrix}, \text{ where } A = [a_{ij}]$$

$$= t^n + c_1 t^{n-1} + c_2 t^{n-2} + \ldots\ldots + c_{n-1} t + c_n, \qquad \ldots (2)$$

where the coefficients $c_1, c_2, \ldots\ldots, c_{n-1}$ and c_n belong to the field F.

We will now use Equations (1) and (2) to prove the Cayley-Hamilton theorem.

Theorem 1 (Cayley-Hamilton): Let $f(t) = t^n + c_1 t^{n-1} + \ldots\ldots + c_{n-1} t + c_n$ be the characteristic polynomial of an $n \times n$ matrix A. Then,

$f(A) = A^n + c_1 A^{n-1} + c_2 A^{n-2} + \ldots + c_{n-1} A + c_n I = 0$

(Note that over here 0 denotes the $n \times n$ zero matrix, and $I = I_n$.)

Proof: We have, by Theorem 3 of Unit 9,

$$\begin{aligned}(tI-A)\,\text{Adj}(tI-A) &= \text{Adj}(tI-A).(tI-A) \\ &= \det(tI-A)I \\ &= f(t)I.\end{aligned}$$

Now Equation (1) above says that

$\text{Adj}(tI-A) = B_1 t^{n-1} + B_2 t^{n-2} + \ldots\ldots + B_n$, where B_k is an $n \times n$ matrix for $k = 1, 2, \ldots, n$.

Thus, we have

$(tI-A)(B_1 t^{n-1} + B_2 t^{n-2} + B_3 t^{n-3} + \ldots\ldots + B_{n-2} t^2 + B_{n-1} t + B_n)$

$= f(t)\, I$

$= It^n + c_1 It^{n-1} + c_2 It^{n-2} + \ldots\ldots + c_{n-2} It^2 + c_{n-1} It + c_n I$, substituting the value of f(t).

Now, comparing the constant term and the coefficients of $t, t^2, \ldots, t^n$ on both sides we get,

$$\begin{aligned} -AB_n &= c_n I \\ B_n - AB_{n-1} &= c_{n-1} I \\ B_{n-1} - AB_{n-2} &= c_{n-2} I \\ &\vdots \\ B_3 - AB_2 &= c_2 I \\ B_2 - AB_1 &= c_1 I \\ B_1 &= I \end{aligned}$$

Pre-multiplying the first equation by I, the second by A, the third by A^2, ... the last by A^n, and adding all these equations, we get

$0 = c_n I + c_{n-1} A + c_{n-2} A^2 + \ldots + c_2 A^{n-2} + c_1 A^{n-1} + A^n = f(A)$

Thus $f(A) = A^n + c_1 A^{n-1} + c_2 A^{n-2} + \ldots + c_{n-1} A + c_n I = 0$, and the Cayley-Hamilton theorem is proved.

This theorem can also be stated as

"Every square matrix satisfies its characteristic polynomial."

Remark 1: You may be tempted to give the following 'quick' proof to Theorem 1:

$f(t) = \det(tI - A)$

$\Rightarrow f(A) = \det(AI - A) = \det(A - A) = \det(0) = 0.$

This proof is false. Why? Well, the left hand side of the above equation, $f(A)$, is an $n \times n$ matrix while the right hand side is the scalar 0, being the value of $\det(0)$.

Now, as usual, we give the analogue of Theorem 1 for linear transformations.

Theorem 2 (Cayley-Hamilton): Let T be a linear transformation on a finite-dimensional vector space V. If $f(t)$ is the characteristic polynomial of T, then $f(T) = 0$.

Proof: Let $\dim V = n$, and let $B = \{v_1, v_2, \ldots, v_n\}$ be a basis of V. In Unit 10 we have observed that

$f(t)$ = the characteristic polynomial of T

= the characteristic polynomial of the matrix $[T]_B$.

Let $[T]_B = A$.

If $f(t) = t^n + c_1t^{n-1} + c_2t^{n-2} + \ldots + c_{n-1}t + c_n$, then, by Theorem 1,

$f(A) = A^n + c_1A^{n-1} + c_2A^{n-2} + \ldots\ldots + c_{n-1}A + c_nI = 0.$

Now, in Theorem 2 of Unit 7 we proved that $[\ ]_B$ is a vector space isomorphism. Thus,

$[f(T)]_B = [T^n + c_1T^{n-1} + c_2T^{n-2} + \ldots + c_{n-1}T + c_nI]_B$

$= [T]_B^n + c_1[T]_B^{n-1} + c_2[T]_B^{n-2} + \ldots + c_{n-1}[T]_B + c_n[I]_B$

$= A^n + c_1A^{n-1} + c_2A^{n-2} + \ldots + c_{n-1}A + c_nI$

$= f(A) = 0$

Again, using the one-one property of $[\ ]_B$, this implies that $f(T) = 0$.

Thus, Theorem 2 is true.

Let us look at some examples now.

Example 1: Verify the Cayley-Hamilton theorem for $A = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}$.

Solution: The characteristic polynomial of A is

$$\begin{vmatrix} t-3 & -2 \\ 1 & t \end{vmatrix} = t^2 - 3t + 2$$

$\therefore$, we want to verify that $A^2 - 3A + 2I = 0$.

$$\text{Now, } A^2 = \begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 7 & 6 \\ -3 & -2 \end{bmatrix}$$

$$\therefore A^2 - 3A + 2I = \begin{bmatrix} 7 & 6 \\ -3 & -2 \end{bmatrix} - \begin{bmatrix} 9 & 6 \\ -3 & 0 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

$\therefore$, the Cayley-Hamilton theorem is true in this case.

E1) Verify the Cayley-Hamilton theorem for A, where A =

a) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ -2 & -2 & 1 \end{bmatrix}$, b) $\begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{bmatrix}$, c) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{bmatrix}$.

We will now use Theorem 1 to prove a result that gives us a method for obtaining the inverse of an invertible matrix.

Theorem 3: Let $f(t) = t^n + c_1t^{n-1} + \ldots\ldots + c_{n-1}t + c_n$ be the characteristic polynomial of an $n \times n$ matrix A. Then A^{-1} exists if $c_n \neq 0$ and, in this case,
$A^{-1} = -c_n^{-1}(A^{n-1} + c_1A^{n-2} + \ldots\ldots + c_{n-1}I)$.

Proof: By Theorem 1,
$f(A) = A^n + c_1A^{n-1} + \ldots + c_{n-1}A + c_nI = 0$.
$\Rightarrow A(A^{n-1} + c_1A^{n-2} + \ldots + c_{n-1}I) = -c_nI$
and $(A^{n-1} + c_1A^{n-2} + \ldots + c_{n-1}I)A = -c_nI$
$\Rightarrow A\{-c_n^{-1}(A^{n-1} + c_1A^{n-2} + \ldots + c_{n-1}I)\} = I$
$= \{-c_n^{-1}(A^{n-1} + c_1A^{n-2} + \ldots + c_{n-1}I)\}A$.
Thus, A is invertible, and
$A^{-1} = -c_n^{-1}(A^{n-1} + c_1A^{n-2} + \ldots + c_{n-1}I)$.

Let us see how Theorem 3 works in practice.

Example 2: Is $A = \begin{bmatrix} 2 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & 1 & 3 \end{bmatrix}$ invertible? If so, find A^{-1}.

Solution: The characteristic polynomial of A, f(t)

$$= \begin{vmatrix} t-2 & -1 & -1 \\ 1 & t-2 & 1 \\ 1 & -1 & t-3 \end{vmatrix} = t^3 - 7t^2 + 19t - 19.$$

Since the constant term of $f(t) = -19 \neq 0$, A is invertible.

Now, by Theorem 1, $f(A) = A^3 - 7A^2 + 19A - 19I = 0$

$\Rightarrow (1/19)A(A^2 - 7A + 19I) = I$

Therefore, $A^{-1} = (1/19)(A^2 - 7A + 19I)$

$$\text{Now, } A^2 = \begin{bmatrix} 2 & 5 & 4 \\ -3 & 2 & -6 \\ -6 & 4 & 7 \end{bmatrix}$$

$$\therefore A^{-1} = 1/19 \begin{bmatrix} 7 & -2 & -3 \\ 4 & 7 & 1 \\ 1 & -3 & 5 \end{bmatrix}$$

To make sure that there has been no error in calculation, multiply this matrix by A. You should get I!

Now try the following exercise.

E E2) For the matrices in E1, obtain A^{-1}, wherever possible.

Now let us look closely at the minimal polynomial.

11.3 MINIMAL POLYNOMIAL

In Unit 6 we defined the minimal polynomial of a linear transformation $T: V \rightarrow V$. We said that it is **the monic polynomial of least degree with coefficients in F, which is satisfied by T.** But, we weren't able to give a method of obtaining the minimal polynomial of T.

In this section we will show that the minimal polynomial divides the characteristic polynomial. Moreover, the roots of the minimal polynomial are the same as those of the characteristic polynomial. Since it is easy to obtain the characteristic polynomial of T, these facts will give us a simple way of finding the minimal polynomial of T.

Let us first recall some properties of the minimal polynomial of T that we gave in Unit 6. Let p(t) be the minimal polynomial of T, then

MP1) p(t) is a monic polynomial with coefficients in F.

MP2) $p(T) = 0$
MP3) If $q(t)$ is a non-zero polynomial over F such that deg q < deg p, then $q(T) \neq 0$.
MP4) If, for some polynomial $g(t)$ over F, $g(T) = 0$, then $p(t) \mid g(t)$. That is, there exists a polynomial $h(t)$ over F such that $g(t) = p(t)h(t)$.

We will now obtain the first link in the relationship between the minimal polynomial and the characteristic polynomial of a linear transformation.

Theorem 4: The minimal polynomial of a linear transformation divides its characteristic polynomial.

Proof: Let the characteristic polynomial and the minimal polynomial of T be $f(t)$ and $p(t)$, respectively. By Theorem 2, $f(T) = 0$. Therefore, by MP4, $p(t)$ divides $f(t)$, as desired.

Before going on to show the full relationship between the minimal and characteristic polynomials, we state (but don't prove!) two theorems that will be used again and again, in this course as well as other courses.

Theorem 5 (Division algorithm for polynomials): Let f and g be two polynomials in t with coefficients in a field F such that $f \neq 0$. Then

a) there exist polynomials q and r with coefficients in F such that $g = fq + r$, where $r = 0$ or deg r < deg f, and

b) if we also have $g = fq_1 + r_1$, with $r_1 = 0$ or deg r_1 < deg f, then $q = q_1$ and $r = r_1$.

An immediate corollary follows.

Corollary: If g is a polynomial over F with $\lambda \in F$ as a root, then $g(t) = (t-\lambda)\, q(t)$, for some polynomial q over F.

Proof: By the division algorithm, taking $f = (t - \lambda)$ we get

$g(t) = (t - \lambda)\, q(t) + r(t)$, ... (1)

with $r = 0$ or deg r < deg $(t - \lambda) = 1$.

If deg r < 1, then r is a constant.

Putting $t = \lambda$ in (1) gives us

$g(\lambda) = r(\lambda) = r$, since r is a constant. But $g(\lambda) = 0$, since λ is a root of g. $\therefore$, $r = 0$. Thus, the only possibility is $r = 0$. Hence, $g(t) = (t - \lambda)\, q(t)$.

And now we come to a very important result that you may be using often, without realising it. The mathematician Gauss gave four proofs of this theorem between 1797 and 1849.

Theorem 6 (Fundamental theorem of algebra): Every non-constant polynomial with complex coefficients has at least one root in C.

In other words, this theorem says that any polynomial $f(t) = \alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0$ (where $\alpha_0, \alpha_1, \ldots, \alpha_n \in C$, $\alpha_n \neq 0$, $n \geq 1$) has at least one root in C.

Remark 2: In Theorem 6, if $\lambda_1 \in C$ is a root of $f(t) = 0$, then, by Theorem 5, $f(t) = (t-\lambda_1)f_1(t)$. Here deg $f_1 = n-1$. If $f_1(t)$ is not constant, then the equation $f_1(t) = 0$ has a root $\lambda_2 \in C$, and $f_1(t) = (t-\lambda_2)f_2(t)$. Consequently, $f(t) = (t-\lambda_1)(t-\lambda_2)\, f_2(t)$. Here deg $f_2 = n-2$. Using the fundamental theorem repeatedly, we get

$f(t) = \alpha_n(t-\lambda_1)(t-\lambda_2)\ldots(t-\lambda_n)$ for some $\lambda_1, \lambda_2, \ldots, \lambda_n$ in C, which are not necessarily distinct. (This process has to stop after n steps since deg f = n.) Thus, all the roots of $f(t)$ belong to C and these are n in number. They may not all be distinct. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the distinct roots, and they are repeated $m_1, m_2, \ldots, m_k$ times, respectively. Then $m_1 + m_2 + \ldots + m_k = n$, and $f(t) = \alpha_n (t - \lambda_1)^{m_1} (t - \lambda_2)^{m_2} \ldots (t - \lambda_k)^{m_k}$.

For example, the polynomial equation $t^3 - it^2 + t - i = 0$ has no real roots, but it has two distinct complex roots, namely, $i\,(= \sqrt{-1})$ and $-i$. And we write $t^3 - it^2 + t - i = (t-i)^2(t+i)$. Here i is repeated twice and $-i$ only occurs once.

We can similarly show that any polynomial $f(t)$ over R can be written as a product of linear polynomials and quadratic polynomials. For example the real polynomial $t^3 - 1 = (t-1)(t^2+t+1)$.

Now we go on to show the second and final link that relates the minimal and characteristic polynomials of $T : V \to V$, where V is a vector space over **F**. Let p(t) be the minimal polynomial of T. We will show that a scalar λ is an eigenvalue of T if and only if λ is a root of p(t). The proof will utilise the following remark.

Remark 3: If λ is an eigenvalue of T, then $Tx = \lambda x$ for some $x \in V$, $x \neq 0$. But $Tx = \lambda x \Rightarrow T^2x = T(Tx) = T(\lambda x) = \lambda Tx = \lambda^2 x$. By induction it is easy to see that $T^k x = \lambda^k x$ for all k. Now, if $g(t) = a_n t^n + a_{n-1} t^{n-1} + \ldots\ldots + a_1 t + a_0$ is a polynomial over F, then $g(T) = a_n T^n + a_{n-1} T^{n-1} + \ldots\ldots + a_1 T + a_0 I$.

This means that

$g(T)x = a_n T^n x + a_{n-1} T^{n-1} x + \ldots\ldots + a_1 Tx + a_0 x$

$= a_n \lambda^n x + a_{n-1} \lambda^{n-1} x + \ldots\ldots + a_1 \lambda x + a_0 x$

$= g(\lambda)\, x$

Thus, λ is an eigenvalue of $T \Rightarrow g(\lambda)$ is an eigenvalue of g(T).

Now for the theorem.

Theorem 7: Let T be a linear transformation on a finite-dimensional vector space V over the field F. Then $\lambda \in F$ is an eigenvalue of T if and only if λ is a root of the minimal polynomial of T. In particular, **the characteristic polynomial and the minimal polynomial of T have the same roots.**

Proof: Let p be the minimal polynomial of T and let $\lambda \in F$. Suppose λ is an eigenvalue of T. Then $Tx = \lambda x$ for some $0 \neq x \in V$. Also, by Remark 3, $p(T)x = p(\lambda)x$. But $p(T) = 0$. Thus, $0 = p(\lambda)x$. Since $x \neq 0$, we get $p(\lambda) = 0$, that is, λ is a root of p(t).

Conversely, let λ be a root of p(t), then $p(\lambda) = 0$ and, by Theorem 5, $p(t) = (t-\lambda)q(t)$, $\deg q < \deg p$, $q \neq 0$. By the property MP3, $\exists\ v \in V$ such that $q(T)v \neq 0$.

Let $x = q(T)v \neq 0$. Then,

$(T - \lambda I)x = (T - \lambda I)\, q(T)v = p(T)v = 0$

$\Rightarrow Tx - \lambda x = 0 \Rightarrow Tx = \lambda x$. Hence, λ is an eigenvalue of T.

So, λ is an eigenvalue of T iff λ is a root of the minimal polynomial of T.

In Unit 10 we have already observed that λ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial of T. Hence, we have shown that both the minimal and characteristic polynomials of T have the same roots, namely, the eigenvalues of T.

Caution: Though the roots of the characteristic polynomial and the minimal polynomial coincide, the two polynomials are not the same, in general.

For example, if the characteristic polynomial of $T : \mathbf{R}^4 \to \mathbf{R}^4$ is $(t+1)^2 (t-2)^2$, then the minimal polynomial could be $(t+1)(t-2)$ or $(t+1)^2 (t-2)$, or $(t+1)(t-2)^2$, or even $(t+1)^2(t-2)^2$, depending on which of these polynomials is satisfied by T.

In general, let $f(t) = (t-\lambda_1)^{n_1} (t-\lambda_2)^{n_2} \ldots (t-\lambda_r)^{n_r}$ be the characteristic polynomial of a linear transformation T, where $\deg f = n$ ($\therefore n_1 + n_2 + \ldots + n_r = n$) and $\lambda_1, \ldots, \lambda_r \in C$ are distinct. Then the minimal polynomial p(t) is given by

$p(t) = (t-\lambda_1)^{m_1} (t-\lambda_2)^{m_2} \ldots (t-\lambda_r)^{m_r}$, where $1 \leq m_i \leq n_i$ for $i = 1, 2, \ldots\ldots, r$.

In case T has n distinct eigenvalues, then

$f(t) = (t-\lambda_1)(t-\lambda_2) \ldots\ldots\ldots (t-\lambda_n)$

and therefore,

$p(t) = (t-\lambda_1)(t-\lambda_2) \cdots\cdots (t-\lambda_n) = f(t)$.

E E3) What can the minimal polynomial of $T : \mathbf{R}^3 \to \mathbf{R}^3$ be if its characteristic polynomial is

a) t^3, b) $t(t-1)(t+2)$?

Analogous to the definition of the minimal polynomial of a linear transformation, we define the minimal polynomial of a matrix.

Definition: **The minimal polynomial of a matrix A over F is the monic polynomial p(t) such that**

i) $p(A) = 0$, and

ii) if $q(t)$ is a non-zero polynomial over **F** such that $\deg q < \deg p$, then $q(A) \neq 0$.

We state two theorems which are analogous to Theorems 4 and 7. Their proofs are also similar to those of Theorems 4 and 7

Theorem 8: The minimal polynomial of a matrix divides its characteristic polynomial.

Theorem 9: The roots of the minimal polynomial and the characteristic polynomial of a matrix are the same, and are the eigenvalues of the matrix.

Let us use these theorems now.

Example 3: Obtain the minimal polynomial of $A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$

Solution: The characteristic polynomial of A =

$$f(t) = \begin{vmatrix} t-5 & 6 & 6 \\ 1 & t-4 & -2 \\ -3 & 6 & t+4 \end{vmatrix} = (t-1)(t-2)^2.$$

Therefore, the minimal polynomial p(t) is either (t–1)(t–2) or $(t-1)(t-2)^2$

Since (A – I) (A – 2I)

$$= \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0,$$

the minimal polynomial, p(t), is (t–1)(t–2).

Example 4 : Find the minimal polynomial of

$$A = \begin{bmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{bmatrix}$$

Solution: The characteristic polynomial of A =

$$f(t) = \begin{vmatrix} t-3 & -1 & 1 \\ -2 & t-2 & 1 \\ -2 & -2 & t \end{vmatrix} = (t-1)(t-2)^2$$

Again, as before, the minimal polynomial p(t) of A is either (t–1) (t–2) or $(t-1)(t-2)^2$. But, in this case,

$$(A-I)(A-2I) = \begin{bmatrix} 2 & 1 & -1 \\ 2 & 1 & -1 \\ 2 & 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 & -1 \\ 2 & 0 & -1 \\ 4 & 0 & -2 \end{bmatrix} \neq 0.$$

Hence, $p(t) \neq (t-1)(t-2)$. Thus, $p(t) = (t-1)(t-2)^2$.

Now, let T be a linear transformation for V to V, and B be a basis of V. Let $A = [T]_B$. If g(t) is any polynomial with coefficients in **F**, then $g(T) = 0$ if and only if $g(A) = 0$. Thus, the minimal polynomial of T is the same as the minimal polynomial of A. So, for example, if $T : \mathbf{R}^3 \to \mathbf{R}^3$ is a linear operator which is represented with respect to the standard basis, by the matrix in Example 3, then its minimal polynomial is (t–1) (t–2).

Example 5: What can the minimal polynomial of $T : \mathbf{R}^4 \to \mathbf{R}^4$ be if the characteristic polynomial of $[T]_B$ is

a) $(t-1)(t^3+1)$, b) $(t^2+1)^2$?

Here, B is the standard basis of $\mathbf{R}^4$.

Solution: a) Now $(t-1)(t^3+1) = (t-1)(t+1)(t^2-t+1)$. This has 4 distinct complex roots, of which only 1 and -1 are real. Since all the roots are distinct this polynomial is also the minimal polynomial of T.

b) $(t^2+1)^2$ has no real roots. It has 2 repeated complex roots, i and $-$i. Now, the minimal polynomial must be a real polynomial that divides the characteristic polynomial. $\therefore$, it can be (t^2+1) or $(t^2+1)^2$.

This example shows you that if the minimal polynomial is a real polynomial, then it need not be a product of linear polynomials only. Of course, over **C** it will always be a product of linear polynomials.

Try the following exercises now.

E E4) Find the minimal polynomial of

a) $A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$

b) $T : \mathbf{R}^3 \to \mathbf{R}^3 : T(x,y,z) = (x+y, y+z, z+x)$

The next exercise involves the concept of the trace of a matrix. If $A = [a_{ij}] \in M_n(F)$, then the trace of A, denoted by Tr(A) is $-$ (coefficient of t^{n-1} in $f_A(t)$).

E5) Let $A = [a_{ij}] \in M_n(F)$. For the matrix A given in E4, show that
$Tr(A)$ = (sum of its eigenvalues)
= (sum of its diagonal elements).

We end the unit by recapitulating what we have done in it.

11.4 SUMMARY

In this unit we have covered the following points.

1) The proof of the Cayley-Hamilton theorem, which says that every square matrix (or linear transformation $T: V \to V$) satisfies its characteristic equation.
2) The use of the Cayley-Hamilton theorem to find the inverse of a matrix.
3) The definition of the minimal polynomial of a matrix.
4) The proof of the fact that the minimal polynomial and the characteristic polynomial of a linear transformation (or matrix) have the same roots. These roots are precisely the eigenvalues of the concerned linear transformation (or matrix).
5) A method for obtaining the minimal polynomial of a linear transformation (or matrix).

11.5 SOLUTIONS/ANSWERS

E 1) a) $$f_A(t) = \begin{vmatrix} t-1 & 0 & 0 \\ -2 & t-3 & 0 \\ 2 & 2 & t-1 \end{vmatrix} = (t-1)^2(t-3)$$

Now, $(A - I)^2 (A - 3I) = 0$. $\therefore$ A satisfies $f_A(t)$.

b) $$f_A(t) = \begin{vmatrix} t & -1 & 0 \\ -3 & t & -1 \\ -1 & 2 & t+1 \end{vmatrix} = t^3 + t^2 - t - 4.$$

$$\text{Now, } A^2 = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{bmatrix}$$

$$\therefore A^3 = \begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{bmatrix}\begin{bmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 2 \\ 8 & -5 & 4 \end{bmatrix}$$

$$\text{Now, } A^3 + A^2 - A - 4I = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & 2 \\ 8 & -5 & 4 \end{bmatrix} + \begin{bmatrix} 3 & 0 & 1 \\ 1 & 1 & -1 \\ -7 & 3 & -1 \end{bmatrix}$$

$$- \begin{bmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 1 & -2 & -1 \end{bmatrix} - \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} = 0.$$

c) $f_A(t) = \begin{vmatrix} t-1 & 0 & -1 \\ 0 & t-3 & -1 \\ -3 & -3 & t-4 \end{vmatrix} = t^3 - 8t^2 + 13t$

$$\text{Now, } A^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 3 & 5 \\ 3 & 12 & 7 \\ 15 & 21 & 22 \end{bmatrix}$$

$$\therefore A^3 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 3 & 1 \\ 3 & 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 3 & 5 \\ 3 & 12 & 7 \\ 15 & 21 & 22 \end{bmatrix} = \begin{bmatrix} 19 & 24 & 27 \\ 24 & 57 & 43 \\ 81 & 129 & 124 \end{bmatrix}$$

$$\therefore A^3 - 8A^2 + 13A = \begin{bmatrix} 19 & 24 & 27 \\ 24 & 57 & 43 \\ 81 & 129 & 124 \end{bmatrix} - \begin{bmatrix} 32 & 24 & 40 \\ 24 & 96 & 56 \\ 120 & 168 & 176 \end{bmatrix}$$

$$+ \begin{bmatrix} 13 & 0 & 13 \\ 0 & 39 & 13 \\ 39 & 39 & 52 \end{bmatrix} = 0.$$

$\therefore$ A satisfies its characteristic polynomial.

E2) a) The constant term of $f_A(t)$ is $-3 \neq 0$. $\therefore$ A is invertible.
Now, $A^3 - 5A^2 + 7A - 3I = 0$.

$$\therefore A^{-1} = \frac{1}{3}(A^2 - 5A + 7I)$$

$$= \frac{1}{3}\left(\begin{bmatrix} 1 & 0 & 0 \\ 8 & 9 & 0 \\ -8 & -8 & 1 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 10 & 15 & 0 \\ -10 & -10 & 5 \end{bmatrix} + \begin{bmatrix} 7 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 7 \end{bmatrix} \right)$$

$$= \frac{1}{3} \begin{bmatrix} 3 & 0 & 0 \\ -2 & 1 & 0 \\ 2 & 2 & 3 \end{bmatrix}.$$

Pre-multiply by A to check that our calculations are right.

b) A is invertible, and $A^{-1} = \frac{1}{4}(A^2 + A - I)$

$$= \frac{1}{4} \begin{bmatrix} 2 & 1 & 1 \\ 4 & 0 & 0 \\ -6 & 1 & -3 \end{bmatrix}$$

c) A is not invertible, by Theorem 3.

E 3) a) The minimal polynomial can be t, t^2 or t^3.

b) The minimal polynomial can only be $t(t-1)(t+2)$.

E 4) a) $f_A(t) = \begin{vmatrix} t & -1 & 0 & -1 \\ -1 & t & -1 & 0 \\ 0 & -1 & t & -1 \\ -1 & 0 & -1 & t \end{vmatrix} = t^2(t-2)(t+2)$

$\therefore$, the minimal polynomial can be $t(t-2)(t+2)$ or $t^2(t-2)(t+2)$.

Now $A(A-2I)(A+2I) = 0$. $\therefore$, $t(t-2)(t+2)$ is the minimal polynomial of A.

b) The matrix of T with respect to the standard basis is

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

Then $f_A(t) = \begin{vmatrix} t-1 & -1 & 0 \\ 0 & t-1 & -1 \\ -1 & 0 & t-1 \end{vmatrix} = t^3 - t^2 - t.$

This has 3 distinct roots: $0, \frac{1+i\sqrt{5}}{2}, \frac{1-i\sqrt{5}}{2}$

$\therefore$, the minimal polynomial is the same as $f_A(t)$.

E5) Sum of diagonal elements = 0.
Sum of eigenvalues = $0 - 2 + 2 = 0$ and $Tr(A) = -$ (coeff. of t^3 in $f_A(t)$) = 0.
$\therefore$ Tr(A) = sum of diagonal elements of A.
= sum of eigenvalues of A.

NOTES

UTTAR PRADESH
RAJARSHI TANDON OPEN UNIVERSITY

UGMM-02
LINEAR ALGEBRA

Block

4

INNER PRODUCTS AND QUADRATIC FORMS

BLOCK 4 INNER PRODUCTS AND QUADRATIC FORMS

This is the last block of this course. It deals with the interesting properties of a special class of vector spaces which are known as inner product spaces. **In this block the only vector spaces we consider will be over R or C.**

In the first unit (Unit 12) we introduce the basic notion of the inner product of two vectors, along with its properties. This product helps us in introducing the well-known geometrical notions of lengths and angles between vectors. We go on to discuss the concept of orthogonality and the solution of the basic problem of the existence of an orthonormal basis in a finite-dimensional inner product space.

The second unit (Unit 13) deals with the problem of characterising linear functionals in inner product spaces. We show that such functionals are always represented as inner products. This further helps us in proving the existence of a unique adjoint for every given operator. Some interesting relations between an operator and its adjoint lead us to define self-adjoint and unitary operators. We also establish some of their properties. Then we introduce you to Hermitian, unitary and orthogonal matrices and the concept of orthogonal similarity.

In Units 14 and 15 we deal with real vector spaces only. The purpose of these two units is to use the methods of linear algebra that you have studied in the course so far, to reduce quadratic forms in $\mathbf{R}^2$ and $\mathbf{R}^3$ into simpler forms. In Unit 15 you will study various conics in detail.

What you study in these units will stand you in good stead in various mathematics courses, particularly in geometry and mechanics.

In case you are interested in knowing more about the material covered in this block, you may look up the books that we have given in the course introduction.

UNIT 12 INNER PRODUCT SPACES

Structure

12.1 INTRODUCTION

So far you have studied many interesting vector spaces over various fields. **In this unit, and the following ones, we will only consider real and complex vector spaces.** In Unit 2 you studied geometrical notions like the length of a vector, the angle between two vectors and the dot product in $\mathbf{R}^2$ or $\mathbf{R}^3$. In this unit we carry these concepts over to a more general setting. We will define a certain special class of vector spaces which open up new and interesting vistas for investigations in mathematics and physics. Hence their study is extremely fruitful as far as the applications of the theory to problems are concerned. This fact will become clear in Units 14 and 15.

Before going further we suggest that you refer to Unit 2 for the definitions and properties of the length and the scalar product of vectors of $\mathbf{R}^2$ or $\mathbf{R}^3$.

Objectives

After reading this unit, you should be able to

- define and give examples of inner product spaces;
- define the norm of a vector and discuss its properties;
- define orthogonal vectors and discuss some properties of sets of orthogonal vectors;
- obtain an orthonormal basis from a given basis of a finite-dimensional inner product space.

12.2 INNER PRODUCT

In this section we start with defining a concept which is the generalisation of the scalar product that you came across in Unit 2. Recall that if (x_1, x_2, x_3) and (y_1, y_2, y_3) are two vectors in $\mathbf{R}^3$, then their scalar product is

$(x_1, x_2, x_3) \cdot (y_1, y_2, y_3) = x_1y_1 + x_2y_2 + x_3y_3$.

We also remind you that given any complex number $z = a + ib$, where $a, b \in \mathbf{R}$, its **complex conjugate** is $\bar{z} = a - ib$.

Further, $z\bar{z} = |z|^2 = a^2 + b^2$, and $\bar{\bar{z}} = z$.

Now we are ready to define an inner product.

Definition: Let V be a vector space over the field F. A map $\langle\ ,\ \rangle: V \times V \to F: \langle\ ,\ \rangle(x, y) = \langle x, y \rangle$ is called an **inner product** (or **scalar product**) over V if it satisfies the following conditions:

F = R or C

IP 1) $\langle x, x \rangle \geq 0 \ \forall x \in V$.

IP 2) $\langle x, x \rangle = 0$ iff $x = \theta$.

IP 3) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all $x, y, z \in V$

IP 4) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ for $\alpha \in F$ and $x, y \in V$

IP 5) $\langle y, x \rangle = \overline{\langle x, y \rangle}$ for all $x, y \in V$. (Here $\overline{\langle x, y \rangle}$ denotes the complex conjugate of the number $\langle x, y \rangle$.)

The scalar $\langle x, y \rangle$ is called the **inner product (or scalar product) of the vector x with the vector y.**

A vector space V over which an inner product has been defined is called an **inner product space**, and is denoted by $(V, \langle\ ,\ \rangle)$.

We make a remark here.

Remark 1: Let $\alpha \in F$. Then $\alpha = \bar{\alpha}$ iff $\alpha \in \mathbf{R}$. So IP5 implies the following statements.

a) $\langle x, x \rangle \in \mathbf{R}\ \forall x \in V$, since $\langle x, x \rangle = \overline{\langle x, x \rangle}$.

b) If $F = \mathbf{R}$, then $\langle x, y \rangle = \langle y, x \rangle\ \forall x, y \in V$.

Now, let us examine a familiar example.

Example 1: Show that $\mathbf{R}^3$ is an inner product space.

Solution: We need to define an inner product on $\mathbf{R}^3$. For this we define $\langle u, v \rangle = u \cdot v\ \forall u, v \in \mathbf{R}^3$ ($\cdot$ denoting the dot product). Then, for $u = (x_1, x_2, x_3)$ and $v = (y_1, y_2, y_3)$, $\langle u, v \rangle = x_1y_1 + x_2y_2 + x_3y_3$. We want to check if $\langle\ ,\ \rangle$ satisfies IP1 – IP5.

i) IP1 is satisfied because $\langle u, u \rangle = x_1^2 + x_2^2 + x_3^2$, which is always non-negative.

ii) Now, $\langle u, u \rangle = 0 \Rightarrow x_1^2 + x_2^2 + x_3^2 = 0 \Rightarrow x_1 = 0,\ x_2 = 0,\ x_3 = 0$, since the sum of positive real numbers is zero if and only if each of them is zero.
$\therefore$, $u = 0$.
Also, if $u = 0$, then $x_1 = 0 = x_2 = x_3$. $\therefore \langle u, u \rangle = 0$.
So, we have shown that IP2 is satisfied by $<\ ,\ >$.

iii) IP3 is satisfied because

$$\langle u + v, w \rangle = (x_1 + y_1)z_1 + (x_2 + y_2)z_2 + (x_3 + y_3)z_3, \text{ where } w = (z_1, z_2, z_3).$$
$$= (x_1z_1 + x_2z_2 + x_3z_3) + (y_1z_1 + y_2z_2 + y_3z_3) = \langle u, w \rangle + \langle v, w \rangle.$$

We suggest that you verify IP4 and IP5. That's what E1 says!

E E1) Check that the inner product on $\mathbf{R}^3$ satisfies IP4 and IP5.

The inner product that we have given in Example 1 can be generalised to the inner product $\langle\ ,\ \rangle$ on $\mathbf{R}^n$ defined by $\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = x_1y_1 + x_2y_2 + \ldots + x_ny_n$. This is called the **standard inner product on $\mathbf{R}^n$.**

Let us consider another example now.

Example 2: Take $F = \mathbf{C}$ and, for $x, y \in \mathbf{C}$, define $\langle x, y \rangle = x\bar{y}$. Show that the map $\langle\ ,\ \rangle: \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ is an inner product.

Solution: IP1 and IP2 are satisfied because, for any complex number x, $x\bar{x} \geq 0$. Also, $x\bar{x} = 0$ if and only if $x = 0$.

To complete the solution you can try E2.

E E2) Show that IP3, IP4 and IP5 are true for Example 2.

In fact, Example 2 can be generalised to C^n for any $n > 0$. We can define the inner product of two arbitrary vectors

$x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbf{C}^n$ by $\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y_i}$. This inner product is called the **standard inner product on C^n**.

The next example deals with a general complex vector space.

Example 3: Let V be a complex vector space of dimension n. Let $B = \{e_1, \ldots, e_n\}$ be a basis of V. Given $x, y \in V$ $\exists$ unique scalars $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbf{C}$ such that

$$x = \sum_{i=1}^{n} a_i e_i,\ y = \sum_{i=1}^{n} b_i e_i.$$

Define $\langle x, y\rangle = \sum_{i=1}^{n} a_i \overline{b_i}$.

Verify that $\langle\ ,\ \rangle$ is an inner product.

Solution: Let $x = \sum_{i=1}^{n} a_i e_i,\ y = \sum_{i=1}^{n} b_i e_i,\ z = \sum_{i=1}^{n} c_i e_i,$

where $a_i, b_i, c_i \in \mathbf{C}\ \forall i = 1, \ldots, n$. Then

$$\langle x, x\rangle = \sum_{i=1}^{n} a_i \overline{a_i} \geq 0 \text{ Also, } \langle x, x\rangle = 0 \Leftrightarrow a_i = 0 \forall i \Leftrightarrow x = 0$$

$$\langle x + y, z\rangle = \sum_{i=1}^{n} (a_i + b_i)\overline{c_i} = \sum_{i=1}^{n} a_i \overline{c_i} + \sum_{i=1}^{n} b_i \overline{c_i} = \langle x, z\rangle + \langle y, z\rangle$$

Also, for any $\alpha \in \mathbf{C}$, $\langle \alpha x, y\rangle = \sum_{i=1}^{n} \alpha a_i \overline{b_i} = \alpha \sum_{i=1}^{n} a_i \overline{b_i} = \alpha\langle x, y\rangle$

Finally, $\overline{\langle y, x\rangle} = \overline{\sum_{i=1}^{n} b_i \overline{a_i}} = \sum_{i=1}^{n} \overline{b_i} a_i = \sum_{i=1}^{n} a_i \overline{b_i} = \langle x, y\rangle$

$\overline{ab} = \bar{a}\bar{b}\ \forall\ a, b \in \mathbf{C}$.

Thus, IP1 – IP5 are satisfied. This proves that $\langle\ ,\ \rangle$ is an inner product on V.

Note that, in Example 3, the inner product depended on the basis of V that we chose. This suggests that an inner product can be defined on any finite-dimensional vector space. In fact, many such products can be defined by choosing different bases in the same vector space.

You may like to try the following exercise now.

E3) Let $X = \{x_1, \ldots, x_n\}$ be a set and V be the set of all functions from X to C. Then, with respect to pointwise addition and scalar multiplication, V is a vector space over C. Now, for any $f, g \in V$, define

$$\langle f, g\rangle = \sum_{i=1}^{n} f(x_i)\overline{g(x_i)}.$$

Show that $(V, \langle\ ,\ \rangle)$ is an inner produce space.

We now state some properties of inner products that immediately follow from IP1 – IP5.

Theorem 1: Let (V, ⟨ , ⟩) be an inner product space. Then, for any $x, y, z \in V$ and $\alpha, \mu \in C$,

a) $\langle \alpha x + \mu y, z\rangle = \alpha \langle x, z\rangle + \mu \langle y, z\rangle$

b) $\langle x, \alpha y + \mu z\rangle = \bar{\alpha} \langle x, y\rangle + \bar{\mu} \langle x, z\rangle$

c) $\langle 0, x\rangle = \langle x, 0\rangle = 0$.

d) $\langle x - y, z\rangle = \langle x, z\rangle - \langle y, z\rangle$

e) $\langle x, z\rangle = \langle y, z\rangle \ \forall z \in V \Rightarrow x = y$.

Proof: We will prove (a) and (c), and leave the rest to you.

a) $\langle \alpha x + \mu y, z\rangle = \langle \alpha x, z\rangle + \langle \mu y, z\rangle$ (by IP3)

$= \alpha \langle x, z\rangle + \mu \langle y, z\rangle$ (by IP4)

c) The vector $0 \in V$ can be written as $0 = 0 \cdot y$ for some $y \in V$.

Thus, $\langle 0, x\rangle = \langle 0 \cdot y, x\rangle = 0 \langle y, x\rangle = 0$.

Then, $\langle x, 0\rangle = \overline{\langle 0, x\rangle} = \bar{0} = 0$.

The proof of this theorem will be complete once you solve E4.

E E4) Prove (b), (d) and (e) of Theorem 1.

We will now discuss the concept of the length of a vector.

12.3 NORM OF A VECTOR

In Unit 2 we defined the length of a vector v in R^2 or R^3 to be $\sqrt{v \cdot v}$. We will extend this definition to the length of a vector in an inner product space.

Definition: If (V, ⟨ , ⟩) is an inner product space and $x \in V$, then the **norm** (or **length**) of the vector x is defined to be $\sqrt{\langle x, x\rangle}$. It is denoted by $\|x\|$.

Hence, $\|x\| = \sqrt{\langle x, x\rangle}$

We make some pertinent remarks here.

Remark 2: a) By IP1, $\langle x, x\rangle \geq 0 \ \forall x \in V$. Thus $\|x\| \geq 0$.
Also, by IP2, $\|x\| = 0$ iff $x = 0$.

b) For any $\alpha \in \mathbf{C}$, we get $\|\alpha x\| = |\alpha| \, \|x\|$,

because $\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{\alpha\overline{\alpha}\langle x, x\rangle} = \sqrt{|\alpha|^2\langle x, x\rangle}$
$= |\alpha|\sqrt{\langle x, x\rangle} = |\alpha| \, \|x\|$.

As in Unit 2, we call $x \in V$ a **unit vector** if $\|x\| = 1$.

E E5) Show that for any $x \in V$, $x \neq 0$, $\dfrac{x}{\|x\|}$ is a unit vector.

E5 leads us to the following definition.

Definition: Given any vector $x \in V$, $x \neq 0$, $\dfrac{x}{\|x\|}$ is the **normalised form** of x.

E5 tells us that the normalised form of a vector is always a unit vector.

We will now prove some results involving norms. The first one is the Cauchy–Schwarz inequality, a generalized version of Theorem 3 in Unit 2. It is very simple, but very important because it allows us to prove many other useful statements.

This inequality was discovered independently by the French mathematician Cauchy, the German mathematician Schwarz and the Russian mathematician Bunyakowski. However, in most of the literature available in English it is ascribed only to Cauchy and Schwarz.

Theorem 2: Let $(V, \langle\ ,\ \rangle)$ be an inner product space and $x, y \in V$.
Then $|\langle x, y\rangle| \leq \|x\| \, \|y\|$.

Proof: If $x = 0$ or $y = 0$, then $|\langle x, y\rangle| = 0 = \|x\| \, \|y\|$.

So, let us assume that $x \neq 0$ and $y \neq 0$. Hence, $\|y\| > 0$.

Let $z = \dfrac{y}{\|y\|}$. Then $z \in V$, and $\|z\| = 1$. Now, for any $\alpha \in \mathbf{F}$, consider the norm of the vector $x - \alpha z \in V$.

$$\|x - \alpha z\|^2 = \langle x - \alpha z, x - \alpha z\rangle$$
$$= \langle x, x\rangle - \alpha\langle z, x\rangle - \overline{\alpha}\langle x, z\rangle + \alpha\overline{\alpha}\langle z, z\rangle, \text{ using Theorem 1.}$$
$$= \|x\|^2 - \overline{\alpha}\langle x, z\rangle - \alpha\overline{\langle x, z\rangle} + \alpha\overline{\alpha}, \text{ since } \langle z, z\rangle = 1.$$
$$= \|x\|^2 - \overline{\alpha}\langle x, z\rangle - \alpha\overline{\langle x, z\rangle} + \alpha\overline{\alpha} + \langle x, z\rangle\overline{\langle x, z\rangle} - \langle x, z\rangle\overline{\langle x, z\rangle},$$
adding and subtracting $\langle x, z\rangle\overline{\langle x, z\rangle}$
$$= \|x\|^2 - |\langle x, z\rangle|^2 + (\langle x, z\rangle - \alpha)\overline{(\langle x, z\rangle - \alpha)}$$
$$= \|x\|^2 - |\langle x, z\rangle|^2 + |\langle x, z\rangle - \alpha|^2$$

Now $\|x - \alpha z\|^2 \geq 0$. This means that $\|x\|^2 - |\langle x, z\rangle|^2 + |\langle x, z\rangle - \alpha|^2 \geq 0 \ \forall \alpha \in \mathbf{F}$.
In particular, if we choose $\alpha = \langle x, z\rangle$, we get
$0 \leq \|x\|^2 - |\langle x, z\rangle|^2$.

Hence, $|\langle x, z\rangle| \leq \|x\|$, that is, $\left|\left\langle x, \dfrac{y}{\|y\|}\right\rangle\right| \leq \|x\|$

$\Leftrightarrow \dfrac{1}{\|y\|}|\langle x, y\rangle| \leq \|x\|$

$\Leftrightarrow |\langle x, y\rangle| \leq \|x\| \, \|y\|$,

which is the required inequality.

Let us see what the Cauchy–Schwarz inequality looks like in some cases.

Example 4: Write the expression for the Cauchy–Schwarz inequality for the vector space given in E3.

Solution: For any $f \in V, \|f\|^2 = \langle f, f\rangle = \sum_{i=1}^{n} |f(x_i)|^2$. Thus, Theorem 2 says that

$$\left|\sum_{i=1}^{n} f(x_i)\overline{g(x_i)}\right| \le \sqrt{\sum_{i=1}^{n} |f(x_i)|^2}\ \sqrt{\sum_{i=1}^{n} |g(x_i)|^2}\ \forall\ f, g \in V.$$

Do try these exercises now.

E6) Write down the expressions for the Cauchy–Schwarz inequality for the spaces given in Examples 1, 2 and 3.

Vectors x and y are called proportional if $\exists\ \alpha \in F$, $\alpha \neq 0$, with $x = \alpha y$.

E7) If $y = \alpha x$, show that $|\langle x, y\rangle| = \|x\|\,\|y\|$.

We come to the next theorem now, which is a generalisation of well-known results of Euclidean geometry.

Theorem 3: If $(V, \langle\ ,\ \rangle)$ is an inner product space and $x, y \in V$, then

a) $\|x+y\| \le \|x\| + \|y\|$ **(Triangle inequality)**

b) $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$ **(Parallelogram law)**

If $z = a + ib \in C$, then
a) the real part of z is a, and is denoted by Re z,
b) $z + \bar{z} = 2\,\text{Re}\, z$
c) $\text{Re}\, z \le |z|$

Proof: a) Now $\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2$.

$$= \|x\|^2 + \langle x, y\rangle + \overline{\langle x, y\rangle} + \|y\|^2.$$

$$= \|x\|^2 + 2\ \text{Re}\langle x, y\rangle + \|y\|^2.$$

$$\le \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2, \text{ since Re } \langle x, y\rangle \le |\langle x, y\rangle|.$$

$$\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 \text{ (by Theorem 2)}$$

$$= (\|x\| + \|y\|)^2$$

Hence, $\|x+y\|^2 \le (\|x\| + \|y\|)^2$

Taking square roots of both sides we obtain
$\|x+y\| \le \|x\| + \|y\|$.

b) To prove the parallelogram law we expand $\|x+y\|^2 + \|x-y\|^2$ to get
$\langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle + \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle$

$= 2(\|x\|^2 + \|y\|^2)$

Thus, (b) is also proved.

The reason (a) is called the triangle inequality is that for any triangle the sum of the lengths of any two sides is greater than or equal to the length of the third side. So, if we consider a triangle in any Euclidean space, two of whose sides are the vectors x and y, then the third side is x + y (see Fig. 1), and hence, $\|x\| + \|y\| \geq \|x + y\|$.

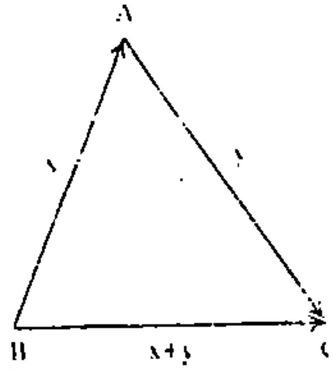


Fig. 1: $\|x + y\| \leq \|x\| + \|y\|$

Similarly, (b) is called the parallelogram law because it generalises the fact that the sum of the squares of the lengths of the diagonals of a parallelogram in Euclidean space is always equal to the sum of the squares of its sides (Fig. 2).

E E8) Show that $|\,\|x\| - \|y\|\,| \leq \|x - y\|$ for $x, y \in (V, \langle\, , \rangle)$.

(Hint: Use the triangle inequality for y and (x − y).)

Let us now discuss a general version of what we did in Sec. 2.5.

12.4 ORTHOGONALITY

In Theorem 2 we showed that $\frac{|\langle x, y\rangle|}{\|x\|\,\|y\|} \leq 1$ for any $x, y \in V$. In Unit 2 (Theorem 2) we have shown that, for non-zero vectors x and y (in $\mathbf{R}^2$ or $\mathbf{R}^3$), $\frac{|\langle x, y\rangle|}{\|x\|\,\|y\|}$ is equal to the magnitude of the cosine of the angle between them. We generalise this concept now.

For any inner product space V and for any non-zero $x, y \in V$, we take $\frac{|\langle x, y\rangle|}{\|x\|\,\|y\|}$ to be the **magnitude of the cosine of the angle between the two vectors x and y.**

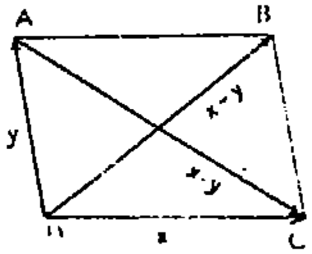


Fig. 2 $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 ...$

So what happens if x and y are perpendicular to each other? We find that $\langle x, y\rangle = 0$. This leads us to the following definition.

Definition: If $(V, \langle\, , \rangle)$ is an inner product space and $x, y \in V$, then x is said to be **orthogonal (or perpendicular)** to y if $\langle x, y\rangle = 0$. This is denoted by $x \perp y$.

For example, $i = (1, 0)$ is orthogonal to $j = (0, 1)$ with respect to the standard inner product in $\mathbf{R}^2$.

We now give some properties involving orthogonality. Their proof is left as an exercise for you.

E9) Using the definitions of inner product and orthogonality, prove the following results for an inner product space V.

a) $0 \perp x \;\forall x \in V$.

b) $x \perp x$ iff $x = 0$, where $x \in V$.

c) $x \perp y \Rightarrow y \perp x$, for $x \in V$.

d) $x \perp y \Rightarrow \alpha x \perp y$ for any $\alpha \in F$, where $x, y \in V$.

Let us consider some examples now.

Example 5: Consider $V = \mathbf{R}^n$. If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are any two vectors of V, we define the inner product of x with y by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

Let $B = \{e_1, \ldots, e_n\}$ be the standard basis of V. Show that $e_i \perp e_j$ when $i \neq j$, $i, j = 1, \ldots, n$. What happens when $i = j$?

Solution: Consider $e_1 = (1, 0, 0, \ldots, 0)$ and $e_2 = (0, 1, 0, \ldots, 0)$. We find that $\langle e_1, e_2 \rangle = 1.0 + 0.1 + 0 + \ldots + 0 = 0$. Hence, $e_1 \perp e_2$. In a similar way, we can show that $e_i \perp e_j$, for $i \neq j$, $i, j = 1, \ldots, n$.

Now let us see what $\langle e_i, e_i \rangle$ is $\forall\, i = 1, \ldots, n$.

$\langle e_1, e_1 \rangle = 1.1 + 0 + \ldots + 0 = 1$.

$\langle e_2, e_2 \rangle = 0 + 1 + 0 + \ldots + 0 = 1$.

Similarly, $\langle e_i, e_i \rangle = 1 \;\forall\, i = 1, \ldots, n$.

Thus, $\| e_i \| = 1 \;\forall\, i = 1, \ldots, n$.

On the lines of Example 5, we can also show that the elements of the standard basis of $\mathbf{C}^n$ are mutually orthogonal and of unit length with respect to the standard inner product.

Try the following exercises now.

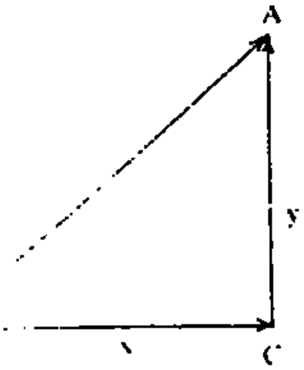


Fig. 3: $\| x+y \|^2 = \| x \|^2 + \| y \|^2$

E E10) For $x, y \in (V, \langle\, ,\, \rangle)$ such that $x \perp y$, show that

$$\| x + y \|^2 = \| x \|^2 + \| y \|^2.$$

(This is the Pythagoras Theorem when $V = \mathbf{R}^2$ (see Fig. 3).)

E E11) Obtain a vector $v = (x, y, z) \in \mathbf{R}^3$ so that v is perpendicular to $(1, 0, 0)$ as well as $(-1, 2, 0)$, with respect to the standard inner product.

We will now define a set of orthogonal vectors.

Definitions: A set $A \subseteq V$ is called **orthogonal** if $x \perp y \;\forall\, x, y \in A$ such that $x \neq y$.
An orthogonal set A is called **orthonormal** if $\| x \| = 1 \;\forall\, x \in A$.
For example, the set B in Example 5 is orthogonal and orthonormal.

By definition, every orthonormal set is orthogonal. But the converse is not true, as the following example tells us.

Example 6: Consider the standard basis $B = \{e_1, \ldots, e_n\}$ of $\mathbf{R}^n$. Show that the set $C = \{2e_1, 2e_2, \ldots, 2e_n\}$ is orthogonal but not orthonormal, with respect to the standard inner product.

Solution: For $i \neq j$, $\langle 2e_i, 2e_j \rangle = 4 \langle e_i, e_j \rangle = 0$. Thus, C is an orthogonal set.

But $\| 2e_i \| = \sqrt{4 \langle e_i, e_i \rangle} = 2 \;\forall\, i = 1, \ldots, n$.

$\therefore$, C is not an orthonormal set.

E12) Let P_n be the real vector space of all real polynomials of degree $\leq n$. We define an inner product on P_n by

$$\left\langle \sum_{i=0}^{n} a_i x^i, \sum_{i=0}^{n} b_i x^i \right\rangle = \sum_{i=0}^{n} a_i b_i.$$

Show that the basis $\{1, x, x^2, \ldots, x^n\}$ of P_n is an orthonormal set.

In the next two theorems we present some properties of an orthogonal set, related to the linear combination of its vectors.

Theorem 4: Let $(V, \langle\, , \rangle)$ be an inner product space and $x, y_1, \ldots, y_n \in V$ such that $x \perp y_i \, \forall \, i = 1, \ldots, n$. Then x is orthogonal to every linear combination of the vectors $y_1, \ldots, y_n$.

Proof: Let $y = \sum_{i=1}^{n} a_i y_i$, where $a_i \in F \, \forall \, i = 1, \ldots, n$.

Then, $y \in V$ and

$$\langle x, y \rangle = \left\langle x, \sum_{i=1}^{n} a_i y_i \right\rangle = \sum_{i=1}^{n} \overline{a_i} \langle x, y_i \rangle = 0, \text{ because } \langle x, y_i \rangle = 0 \, \forall \, i.$$

This shows that $x \perp y$.

Theorem 5: Let $(V, \langle\, , \rangle)$ be an inner product space and $A = \{x_1, \ldots, x_n\} \subseteq V$ be an orthogonal set. Then, for any $a_i \in F$ $(i = 1, \ldots, n)$, we have

$$\left\| \sum_{i=1}^{n} a_i x_i \right\|^2 = \sum_{i=1}^{n} |a_i|^2 \| x_i \|^2.$$

Proof: Our hypothesis says that $\langle x_i, x_j \rangle = 0$ if $i \neq j$. Consider $y = \sum_{i=1}^{n} a_i x_i$.

$$\| y \|^2 = \langle y, y \rangle = \left\langle \sum_{i=1}^{n} a_i x_i, \sum_{j=1}^{n} a_j x_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle a_i x_i, a_j x_j \rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i \overline{a_j} \langle x_i, x_j \rangle$$

$$= \sum_{i=1}^{n} a_i \overline{a_i} \langle x_i, x_i \rangle, \text{ since } \langle x_i, x_j \rangle = 0 \text{ for } i \neq j$$

$$= \sum_{i=1}^{n} |a_i|^2 \| x_i \|^2$$

This proves the result.

Note: If $a_i = 1 \, \forall \, i$, in Theorem 5, we get

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \sum_{i=1}^{n} \| x_i \|^2$$

This is a generalised form of what we gave in E10.

We now give an important result, which is actually a corollary to Theorem 5.

Theorem 6: Let A be an orthogonal set of non-zero vectors of an inner product space V. Then A is a linearly independent set.

Proof: To show that A is linearly independent we will have to prove that any finite subset $\{x_1, \ldots, x_n\}$ of vectors of A is linearly independent. For this, assume that $y = \sum_{i=1}^{n} a_i x_i = 0$.

Then $\|y\|^2 = 0 \Rightarrow \sum_{i=1}^{n} |a_i|^2 \|x_i\|^2 = 0 \Rightarrow |a_i|^2 \|x_i\|^2 = 0 \ \forall\, i$.

$|a_i|^2 = 0$ for $i = 1, \ldots, n$ [illegible] $\|x_i\|^2 \neq 0$ for any i

$a_i = 0$ for [illegible] $, n$

[illegible] is linearly independent. Hence, A is linearly [illegible]

[illegible] that any orthogonal set is linearly [illegible] orthogonal set in a vector space V of dimension n [illegible] elements. So for example, any orthogonal subset of $\mathbf{R}^3$ can have [illegible]

We shall use Theorem 6 as a stepping stone towards showing that any inner product space has an orthonormal set as a basis. But first, some definitions and remarks.

Definition: A basis of an inner product space is called an **orthonormal basis** if its elements form an orthonormal set.

For example, the standard basis of $\mathbf{R}^n$ is an orthonormal basis (Example 5).

Now, a small exercise.

E E13) Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for a real inner product space V. Let

$$x = \sum_{i=1}^{n} x_i e_i \ \text{ and } \ y = \sum_{i=1}^{n} y_i e_i \ \text{ be elements of V. Show that } \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

We make a few observations now.

Remark 3: a) If [illegible] V is orthogonal, then the set

$B = \left\{ \frac{x}{\|x\|} \right\}$ [illegible] and $x \neq 0$ [illegible] orthonormal. For example, consider $\mathbf{R}^2$ with the dot product. Let $v = ($[illegible]$)$ and $w = ($[illegible]$)$. Then [illegible] $= 0$. Thus, $v \perp w$. Therefore,

$\left\{ \frac{v}{\|v\|}, \frac{w}{\|w\|} \right\} =$ [illegible] is an orthonormal set in $\mathbf{R}^2$. In fact, this is a basis of $\mathbf{R}^2$ since $\{v, w\}$ is a linearly independent set and $\dim_{\mathbf{R}} \mathbf{R}^2 = 2$.

b) For any $0 \neq x \in V$, $\left\{ \frac{x}{\|x\|} \right\}$ can be regarded as an orthonormal set in V.

We now state the theorem that tells us of the existence of an orthonormal basis. Its proof consists of a method called the **Gram–Schmidt orthogonalisation process.**

Theorem 7: Let $(V, \langle\ ,\ \rangle)$ be a non-zero inner product space of dimension n. Then V has an orthonormal basis.

Proof: We shall first show that V has an orthogonal basis, and then obtain an orthonormal basis.

[illegible] be a basis of V [illegible] we shall obtain [illegible] orthogonal basis [illegible]

Take $w_1 = v_1$. Define $w_2 = v_2 - \frac{\langle v_2, w_1\rangle}{\langle w_1, w_1\rangle} w_1$. Then $w_2 = v_2 - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle} v_1$,

and

$\langle w_2, v_1\rangle = \langle v_2, v_1\rangle - \frac{\langle v_2, v_1\rangle}{\langle v_1, v_1\rangle}\langle v_1, v_1\rangle = 0$. That is, $\langle w_2, w_1\rangle = 0$. Further, $v_2 = c_1 v_1 + w_2$,

where $c_1 = \frac{\langle v_2, w_1\rangle}{\langle w_1, w_1\rangle} \in F$

Define $w_3 = v_3 - \frac{\langle v_3, w_2\rangle}{\langle w_2, w_2\rangle} w_2 - \frac{\langle v_3, w_1\rangle}{\langle w_1, w_1\rangle} w_1$. Then $\langle w_3, w_2\rangle = 0 = \langle w_3, w_1\rangle$. Also,

$v_3 = c_1 w_1 + c_2 w_2 + w_3$, where $c_1, c_2 \in$ **F**. Continuing in this manner, we can define

$w_{m+1} = v_{m+1} - c_1 w_1 - c_2 w_2 - \cdots - c_m w_m$, where $c_i = \frac{\langle v_{m+1}, w_i\rangle}{\langle w_i, w_i\rangle} \in$ F.

$$\Rightarrow v_{m+1} = c_1 w_1 + c_2 w_2 + \cdots + c_m w_m + w_{m+1} \text{ for } m = 0, \ldots, n-1.$$

This way we obtain an orthogonal set of vectors $\{w_1, w_2, \ldots, w_n\}$, such that the v_i's are a linear combination of the w_i's. By Theorem 6 this set is linearly independent, and hence forms a basis of V.

From this basis, we immediately obtain an orthonormal basis of V by using Remark 3. Thus, $\left\{\frac{w_1}{\|w_1\|}, \ldots, \frac{w_n}{\|w_n\|}\right\}$ is an orthonormal basis of V.

Note: The same process can be used to show that:

If $(V, \langle\ ,\ \rangle)$ is an inner product space and $Y = \{y_1, \ldots, y_n\}$ a set of linearly independent vectors of V, then an orthonormal set $X = \{x_1, x_2, \ldots, x_n\}$ can be obtained from Y such that the linear spans (ref. Unit 3) of X and Y coincide.

Let us see how the Gram–Schmidt process works in a few cases.

Example 7: Obtain an orthonormal basis for $\mathbf{P}_2$, the space of all real polynomials of degree at most 2, the inner product being defined by

$$\langle p_1, p_2\rangle = \int_0^1 p_1(t)\, p_2(t)\, dt.$$

See Block 3 of the Calculus course for definite integrals.

Solution: $\{1, t, t^2\}$ is a basis for $\mathbf{P}_2$. From this we will obtain an orthogonal basis $\{w_1, w_2, w_3\}$. Now $w_1 = 1$ and $\langle w_1, w_1\rangle = \int_0^1 dt = 1$.

$w_2 = t - \frac{\langle t, w_1\rangle}{\langle w_1, w_1\rangle} w_1$. Now $\langle t, w_1\rangle = \int_0^1 t\,dt = \frac{t^2}{2}\Big|_0^1 = \frac{1}{2}$. $\therefore$ $w_2 = t - \frac{1}{2}$

$\therefore \langle w_2, w_2\rangle = \int_0^1\left(t - \frac{1}{2}\right)^2 dt = \frac{1}{12}$.

$$w_3 = t^2 - \frac{\langle t^2, w_2\rangle}{\langle w_2, w_2\rangle} w_2 - \frac{\langle t^2, w_1\rangle}{\langle w_1, w_1\rangle} w_1$$

$$= t^2 - 12\left\{\frac{1}{12}\left(t - \frac{1}{12}\right)\right\} - \frac{1}{3} = t^2 - t + \frac{1}{6}. \text{ Also } \langle w_3, w_3\rangle = \int_0^1 \left(t^2 - t + \frac{1}{6}\right)^2 dt = \frac{1}{180}.$$

Thus, the orthonormal basis is $\left\{\frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}, \frac{w_3}{\|w_3\|}\right\} = \left\{1, \sqrt{12}\left(t - \frac{1}{2}\right), \sqrt{180}\left(t^2 - t + \frac{1}{6}\right)\right\}$.

Here's an exercise.

E E14) Obtain an orthonormal basis, with respect to the standard inner product, for

a) the subspace of $\mathbf{R}^3$ generated by (1, 0, 3) and (2, 1, 1).

b) the subspace of $\mathbf{R}^4$ generated by (1, 0, 2, 0) and (1, 2, 3, 1).

We will now prove a theorem that leads us to an important inequality, which is used for studying Fourier coefficients.

Theorem 8: Let $(V, \langle\, ,\rangle)$ be an inner product space and $A = \{x_1, \ldots, x_n\}$ be an orthonormal set in V. Then, for any $y \in V$,

$$\left\| y - \sum_{i=1}^{n} \langle y, x_i\rangle x_i \right\|^2 = \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2.$$

Proof: Let $x = \sum_{i=1}^{n} a_i x_i$ ($a_i \in F$) be any linear combination of the elements of A.

Then $\|y-x\|^2 = \langle y-x, y-x\rangle = \|y\|^2 - \langle y, x\rangle - \langle x, y\rangle + \|x\|^2$

$$= \|y\|^2 - \left\langle y, \sum_{i=1}^{n} a_i x_i \right\rangle - \left\langle \sum_{i=1}^{n} a_i x_i, y \right\rangle + \|x\|^2$$

$$= \|y\|^2 - \left\langle y, \sum_{i=1}^{n} a_i x_i \right\rangle - \sum_{i=1}^{n} \langle a_i x_i, y\rangle + \sum_{i=1}^{n} |a_i|^2 \|x_i\|^2, \text{ since } \langle x_i, x_j\rangle = 0 \text{ for } i \neq j.$$

As $\|x_i\|^2 = 1 \ \forall\, i$, it follows that

$$\|y-x\|^2 = \|y\|^2 - \sum_{i=1}^{n} \overline{a_i}\langle y, x_i\rangle - \sum_{i=1}^{n} a_i \langle x_i, y\rangle + \sum_{i=1}^{n} a_i \overline{a_i}$$

$$= \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 + \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 - \sum_{i=1}^{n} \overline{a_i}\langle y, x_i\rangle - \sum_{i=1}^{n} a_i \langle x_i, y\rangle + \sum_{i=1}^{n} a_i \overline{a_i}$$

$$= \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 + \sum_{i=1}^{n} \langle y, x_i\rangle \overline{\langle y, x_i\rangle} - \sum_{i=1}^{n} \overline{a_i}\langle y, x_i\rangle - \sum_{i=1}^{n} a_i \overline{\langle y, x_i\rangle} + \sum_{i=1}^{n} a_i \overline{a_i}$$

$$= \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 + \sum_{i=1}^{n} \langle y, x_i\rangle\{\overline{\langle y, x_i\rangle} - \overline{a_i}\} - \sum_{i=1}^{n} a_i \{\overline{\langle y, x_i\rangle} - \overline{a_i}\}$$

$$= \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 + \sum_{i=1}^{n} \{\langle y, x_i\rangle - a_i\}\{\overline{\langle y, x_i\rangle - a_i}\}$$

$$= \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2 + \sum_{i=1}^{n} |\langle y, x_i\rangle - a_i|^2$$

This is true for any $a_i \in F$. Now choose $a_i = \langle y, x_i\rangle \; \forall i, \ldots, n$. Then we get

$$\left\| y - \sum_{i=1}^{n} \langle y, x_i\rangle x_i \right\|^2 = \|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i\rangle|^2,$$

which is the desired result.

And now we come to a corollary of Theorem 8, known as **Bessel's inequality**. It is named after the German astronomer, Friedrich Wilhelm Bessel (1784–1846).

Corollary: Let $A = \{x_1, \ldots, x_n\}$ be any orthonormal set in $(V, \langle\, , \rangle)$. Then, for any $y \in V$,

$$\sum_{i=1}^{n} |\langle y, x_i\rangle|^2 \le \|y\|^2$$

E E15) Prove the corollary given above.

We end the unit by summarising what we have covered in it.

12.5 SUMMARY

In this unit we have discussed the following points. We have

1. defined and given examples of inner product spaces.
2. defined the norm of a vector.
3. proved the Cauchy–Schwarz inequality.
4. defined an orthogonal and an orthonormal set of vectors.
5. shown that every finite-dimensional inner product space has an orthonormal basis, using the Gram–Schmidt orthogonalisation process.
6. proved Bessel's inequality.

12.6 SOLUTIONS/ANSWERS

E1) For $\alpha \in \mathbf{R}$ and $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathbf{R}^3$,

$\langle \alpha(x_1, x_2, x_3), (y_1, y_2, y_3)\rangle = (\alpha x_1, \alpha x_2, \alpha x_3) \cdot (y_1, y_2, y_3)$

$= \alpha x_1 y_1 + \alpha x_2 y_2 + \alpha x_3 y_3 = \alpha(x_1 y_1 + x_2 y_2 + x_3 y_3)$

$= \alpha\langle (x_1, x_2, x_3), (y_1, y_2, y_3)\rangle$.

$\therefore$ IP4 is satisfied.

Also, for any $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ in $\mathbf{R}^3$,

$\langle x, y\rangle = x_1 y_1 + x_2 y_2 + x_3 y_3 = y_1 x_1 + y_2 x_2 + y_3 x_3 = \langle y, x\rangle$.

$\therefore$ IP5 is satisfied.

E2) For $x, y, z \in C$ and $\alpha \in C$ we have

$\langle x+y, z\rangle = (x+y)\bar{z} = x\bar{z} + y\bar{z} = \langle x, z\rangle + \langle y, z\rangle$,

$\langle \alpha x, y\rangle = (\alpha x)\bar{y} = \alpha(x\bar{y}) = \alpha\langle x, y\rangle$,

$\overline{\langle x, y\rangle} = \overline{x\bar{y}} = \bar{x}y = y\bar{x} = \langle y, x\rangle$.

$\therefore \langle\ ,\ \rangle$ satisfies IP3, IP4 and IP5.

E3) Let $f, g, h \in V$ and $\alpha \in C$. Then

$$\langle f, f\rangle = \sum_{i=1}^{n} f(x_i)\overline{f(x_i)} = \sum_{i=1}^{n} |f(x_i)|^2 \geq 0.$$

$\langle f, f\rangle = 0 \Leftrightarrow f(x_i) = 0\ \forall i = 1, \ldots, n$

$\Leftrightarrow$ f is the zero function.

$$\begin{aligned}\langle f+g, h\rangle &= \sum_{i=1}^{n}(f+g)(x_i)\overline{h(x_i)}\\ &= \sum_{i=1}^{n}(f(x_i)+g(x_i))\overline{h(x_i)}\\ &= \sum_{i=1}^{n} f(x_i)\overline{h(x_i)} + \sum_{i=1}^{n} g(x_i)\overline{h(x_i)}\\ &= \langle f, h\rangle + \langle g, h\rangle.\end{aligned}$$

$$\begin{aligned}\langle \alpha f, g\rangle &= \sum_{i=1}^{n} (\alpha f)(x_i)\overline{g(x_i)} = \sum_{i=1}^{n} \alpha f(x_i)\overline{g(x_i)}\\ &= \alpha\sum_{i=1}^{n} f(x_i)\overline{g(x_i)} = \alpha\langle f, g\rangle\end{aligned}$$

$$\begin{aligned}\overline{\langle f, g\rangle} &= \overline{\sum_{i=1}^{n} f(x_i)\overline{g(x_i)}} = \sum_{i=1}^{n} \overline{f(x_i)}g(x_i)\\ &= \sum_{i=1}^{n} g(x_i)\overline{f(x_i)} = \langle g, f\rangle.\end{aligned}$$

$\therefore$ $(V, \langle\ ,\ \rangle)$ is an inner product space.

E4) b) $\langle x, \alpha y + \mu z\rangle = \overline{\langle \alpha y + \mu z, x\rangle}$, by IP5

$= \overline{\alpha\langle y, x\rangle + \mu\langle z, x\rangle}$, by Theorem 1(a).

$= \bar{\alpha}\overline{\langle y, x\rangle} + \bar{\mu}\overline{\langle z, x\rangle}$

$= \bar{\alpha}\langle x, y\rangle + \bar{\mu}\langle x, z\rangle$, by IP5.

$\therefore$ (b) is proved.

d) $\langle x-y, z\rangle = \langle x+(-1)y, z\rangle = \langle x, z\rangle + (-1)\langle y, z\rangle$, by Theorem 1(a).

$= \langle x, z\rangle - \langle y, z\rangle$.

e) $\langle x, z\rangle = \langle y, z\rangle\ \forall z \in V$

$\Rightarrow \langle x-y, z\rangle = 0\ \forall z \in V$, by (d) above.

$\Rightarrow \langle x-y, x-y\rangle = 0$, taking $z = x-y$, in particular.

$\Rightarrow x-y = 0$, by IP2.

$\Rightarrow x = y$.

E5) Let $u = \frac{x}{\|x\|}$. Then $\langle u, u\rangle = \left\langle \frac{x}{\|x\|}, \frac{x}{\|x\|}\right\rangle = \frac{1}{\|x\|^2}\langle x, x\rangle$

$= \frac{1}{\|x\|^2}\|x\|^2 = 1$.

$\therefore \|u\| = \sqrt{\langle u, u\rangle} = 1$.

E6) In the situation of Example 1 we get

$|u \cdot v| \leq \|u\|\|v\|$ for $u, v \in R^3$. Thus,

$|x_1y_1 + x_2y_2 + x_3y_3| \leq \sqrt{x_1^2 + x_2^2 + x_3^2}\sqrt{y_1^2 + y_2^2 + y_3^2}$

$\forall (x_1, x_2, x_3), (y_1, y_2, y_3) \in R^3$.

In the situation of Example 2 we get

$|x\bar{y}| \leq |x||y|\ \forall x, y \in C$.

Theorem 2 and Example 3 gives us

$$\left|\sum_{i=1}^{n} a_i \overline{b_i}\right| \le \sqrt{\sum_{i=1}^{n}|a_i|^2}\,\sqrt{\sum_{i=1}^{n}|b_i|^2}\,.$$

where $\sum_{i=1}^{n} a_i e_i, \sum_{i=1}^{n} b_i e_i$ are elements of V.

E7) $\|y\| = \sqrt{\langle \alpha x, \alpha x\rangle} = \sqrt{|\alpha|^2 \langle x, x\rangle} = |\alpha|\,\|x\|.$

$\therefore \|x\|\|y\| = |\alpha|\,\|x\|^2 = |\overline{\alpha}|\langle x, x\rangle = |\overline{\alpha}\langle x, x\rangle| = |\langle x, \alpha x\rangle| = |\langle x, y\rangle|.$

E8) $\|y + (x-y)\| \le \|y\| + \|x-y\|$

$\Rightarrow \|x\| \le \|y\| + \|x-y\|.$

$\Rightarrow \|x\| - \|y\| \le \|x-y\|.$

Similarly, $\|y\| - \|x\| \le \|y-x\| = \|x-y\|$, since $\|x\| = \|-x\|$.

$\therefore \big|\,\|x\| - \|y\|\,\big| \le \|x-y\|$, since $|\alpha| = \alpha$ or $-\alpha$ for any $\alpha \in \mathbb{R}$

E9) a) Use Theorem 1(c).

b) Since $\langle x, x\rangle = 0 \Leftrightarrow x = 0$, (b) is true.

c) $x \perp y \Rightarrow \langle x, y\rangle = 0 \Rightarrow \overline{\langle y, x\rangle} = 0 \Rightarrow \langle y, x\rangle = 0$

$\Rightarrow y \perp x.$

d) $x \perp y \Rightarrow \langle x, y\rangle = 0 \Rightarrow \alpha\langle x, y\rangle = 0 \;\forall\, \alpha \in F$

$\Rightarrow \langle \alpha x, y\rangle = 0 \;\forall\, \alpha \in F \Rightarrow \alpha x \perp y \;\forall\, \alpha \in F.$

E10) If $x \perp y$, then $\langle x, y\rangle = 0$. Hence, $\|x+y\|^2 = \langle x, x\rangle + \langle y, y\rangle = \|x\|^2 + \|y\|^2$.

E11) $v \perp (1, 0, 0) \Rightarrow x \cdot 1 + y \cdot 0 + z \cdot 0 = 0 \Rightarrow x = 0$

$v \perp (-1, 2, 0) \Rightarrow x \cdot (-1) + y \cdot 2 + z \cdot 0 = 0 \Rightarrow -x + 2y = 0.$

So we get $x = 0, y = 0$. Thus, v is of the form $(0, 0, z)$ for $z \in \mathbb{R}$.

E12) For $i \ne j$, $\langle x^i, x^j\rangle = 1.0 + 0.1 = 0$.

Also, $\forall\, i = 0, \ldots, n$, $\langle x^i, x^i\rangle = 1.1 = 1$.

$\therefore$, the given set is orthonormal.

E13) $\langle x, y\rangle = \left\langle \sum_i x_i e_i, \sum_j y_j e_j \right\rangle = \sum_i \sum_j x_i \overline{y_j} \langle e_i, e_j\rangle$

$= \sum_i x_i \overline{y_i}$, since $\langle e_i, e_i\rangle = 1 \;\forall\, i = 1, \ldots, n$ and

$\langle e_i, e_j\rangle = 0$ for $i \ne j$.

E14) a) Here $v_1 = (1, 0, 3)$, $v_2 = (2, 1, 1)$.

We want the set $\left\{\frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}\right\}$, where $w_1 = v_1$ and

$$w_2 = v_2 - \frac{\langle v_2, w_1\rangle}{\langle w_1, w_1\rangle} w_1$$

Now, $\langle v_2, w_1\rangle = \langle v_2, v_1\rangle = 2 + 0 + 3 = 5.$

Also $\langle w_1, w_1\rangle = \langle v_1, v_1\rangle = 10$, so that $\|w_1\| = \sqrt{10}$.

$$\therefore w_2 = (2, 1, 1) - \frac{5}{10}(1, 0, 3) = \left(\frac{3}{2}, 1, \frac{-1}{2}\right)$$

$$\therefore \|w_2\| = \sqrt{\frac{9}{4} + 1 + \frac{1}{4}} = \sqrt{\frac{7}{2}}$$

$\therefore \left\{\frac{1}{\sqrt{10}}(1, 0, 3), \sqrt{\frac{2}{7}}\left(\frac{3}{2}, 1, \frac{-1}{2}\right)\right\}$ is the required orthonormal basis.

b) $w_1 = (1, 0, 2, 0)$

$$w_2 = (1, 2, 3, 1) - \frac{7}{5}(1, 0, 2, 0) = \left(-\frac{2}{5}, 2, \frac{1}{5}, 1\right),$$

$\|w_1\| = \sqrt{5}$, $\|w_2\| = \sqrt{\frac{26}{5}}$

Then $\left\{\frac{w_1}{\|w_1\|}, \frac{w_2}{\|w_2\|}\right\}$ is the required basis.

E15) Theorem 8 says that

$$\|y\|^2 - \sum_{i=1}^{n} |\langle y, x_i \rangle|^2 \geq 0.$$

UNIT 13 HERMITIAN AND UNITARY OPERATORS

Structure

13.1 INTRODUCTION

In the preceding unit we discussed general properties of inner product spaces. In this unit we will show that we can precisely determine the nature of linear functionals defined over inner product spaces.

We, then, discuss the adjoint of an operator. The behaviour of this adjoint leads us to the concepts of self-adjoint operators and unitary operators. As usual, we will discuss their matrix analogues also. This will entail studying the definitions and properties of Hermitian, unitary and orthogonal matrices.

Regarding the notation in this unit, F will always denote **R** or **C**. And, unless otherwise mentioned, the inner product on $\mathbf{R}^n$ or $\mathbf{C}^n$ will be the standard inner product (ref. Sec. 12.2). Also, if T is a function acting on x, then we will often write Tx for T(x), for our convenience.

Before reading this unit we advise you to look at Unit 6 for the definitions of a linear functional and a dual space.

Objectives

After going through this unit, you should be able to

- represent a linear functional on an inner product space as an inner product with a unique vector;
- prove the existence of a unique adjoint of any given linear operator on an inner product space;
- identify self-adjoint, Hermitian, unitary and orthogonal linear operators;
- establish the relationship between self-adjoint (or unitary) operators and Hermitian (or unitary) matrices.
- prove and use the fact that a matrix is unitary iff its rows (or columns) form an orthogonal set of vectors;
- use the fact that any real symmetric matrix is orthogonally similar to a diagonal matrix.

13.2 LINEAR FUNCTIONALS OF INNER PRODUCT SPACES

If V is a non-zero inner product space over F, then $\exists\, \mathbf{0} \neq x \in V$. Consider the linear functional f on V defined by

$f(v) = \langle v, x\rangle \ \forall\, v \in V$.

Then $f(x) \neq 0$, since $x \neq \mathbf{0}$. Therefore, $f \neq 0$. Also, $f \in V^*$. Therefore, $V^* \neq \{0\}$. But, what do the elements of V^* look like?

Before going into the detailed study of such functionals let us consider an example.

Example 1: Consider $V = \mathbf{R}^2$. Take $y = (1, 2) \in \mathbf{R}^2$ and define, for any $x = (x_1, x_2) \in \mathbf{R}^2$, $f : \mathbf{R}^2 \to \mathbf{R}$ by $f(x) = \langle x, y\rangle = x_1 + 2x_2$. Show that f is a linear functional on $\mathbf{R}^2$

Solution: Firstly, $f[(x_1, x_2) + (y_1, y_2)] = f(x_1, x_2) + f(y_1, y_2) \; \forall (x_1, x_2), (y_1, y_2) \in \mathbf{R}^2$.
Also, for any $a \in \mathbf{R}$, $f(a(x_1, x_2)) = af(x_1, x_2) \; \forall (x_1, x_2) \in \mathbf{R}^2$. Therefore, f is a linear functional on $\mathbf{R}^2$.

Try the following exercise on the same lines as Example 1.

E E1) Fix $y \in \mathbf{R}^2$. Show that the function
$f_y : \mathbf{R}^2 \to \mathbf{R} : f_y(x) = \langle x, y\rangle$ is a linear functional on $\mathbf{R}^2$.

Let us now consider any inner product space $(V, \langle\, ,\rangle)$. We choose a vector $z \in V$ and fix it. With the help of this vector we can obtain a linear functional $f \in V^* = L(V, F)$ in the following way:

define $f : V \to F$ by $f(x) = \langle x, z\rangle \; \forall x \in V$. Clearly f is a well-defined map, and

$f(x + y) = \langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$
$= f(x) + f(y)$.

Also $f(\alpha x) = \langle \alpha x, z\rangle = \alpha\langle x, z\rangle = \alpha f(x)$ for any $\alpha \in F$.
Hence, f is a linear functional on V. (To show the relationship of f with z, we sometimes denote f by f_z)

Thus, we have succeeded in proving the following result.

Theorem 1: If $(V, \langle\, ,\rangle)$ is an inner product space over F(F = R or C) and z is a given vector of V, then the map
$f_z : V \to F : f_z(x) = \langle x, z\rangle$,
is a linear functional on V.

Theorem 1 is true for any finite-dimensional or infinite-dimensional inner product space. What is interesting about finite-dimensional inner product spaces is that the converse of this result is also true. We now proceed to state and prove it.

Theorem 2: If $(V, \langle\, ,\rangle)$ is an inner product space over F with dimension n, and f is a linear functional defined on V, then $\exists$ a unique element z in V such that $f(x) = \langle x, z\rangle$ for $x \in V$, that is, $f = f_z$.

Proof: As dim V = n, it follows from Unit 12 (Theorem 7) that there exists a finite orthonormal basis for V. Let this basis be $B = \{e_1, e_2, \ldots, e_n\}$. Then

$$\langle e_i, e_j\rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

Let $f(e_i) = a_i$ $(i = 1, \ldots, n)$.

Now, any $x \in V$ can be written as $x = \sum_{i=1}^{n} b_i e_i$, $b_i \in F$.

$$\text{Then } f(x) = f\left(\sum_{i=1}^{n} b_i e_i\right) = \sum_{i=1}^{n} b_i f(e_i) = \sum_{i=1}^{n} b_i a_i \qquad \ldots\ldots (1)$$

Now consider the vector $z \in V$ such that $z = \sum_{i=1}^{n} \overline{a_i}\, e_i$.

As each a_i is known to us z is a known vector of V. Also,

$$\langle x, z\rangle = \left\langle \sum_{i=1}^{n} b_i e_i, \sum_{j=1}^{n} \overline{a_j} e_j \right\rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \langle b_i e_i, \overline{a_j} e_j \rangle$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} b_i a_j \langle e_i, e_j \rangle$$

$$= \sum_{i=1}^{n} b_i a_i, \text{ since B is an orthonormal set.}$$

$= f(x)$, from (1) above.

Thus, $f(x) = \langle x, z \rangle \; \forall x \in V$.

Suppose there also exists $z_1 \in V$ such that $f(x) = \langle x, z_1 \rangle \; \forall x \in V$.

Then, $\langle x, z \rangle - \langle x, z_1 \rangle = 0$ for all $x \in V$, i.e.

$\langle x, z - z_1 \rangle = 0$ for all $x \in V$.

Hence, by Unit 12 (Theorem 1), we obtain $z - z_1 = 0$, i.e., $z = z_1$.

Thus, there exists a unique $z \in V$ such that

$f(x) = \langle x, z \rangle \; \forall x \in V$.

We can also represent f in Theorem 2 by $f = \langle \;, z \rangle$. Thus, in Example 1, $f = \langle \;, (1, 2) \rangle$.

See if Theorem 2 can help you in solving the following exercise.

E E2) Define $f : C^3 \to C$ by $f(z_1, z_2, z_3) = \dfrac{(z_1 + z_2 + z_3)}{3}$.
Find the vector $y \in C^3$ such that $f = \langle \;, y \rangle$.

Let us now use linear functionals to define the adjoint of a linear transformation from V to V.

13.3 ADJOINT OF AN OPERATOR

In this section we will obtain a linear transformation from V to V, which corresponds to a given linear operator $T : V \to V$.

Let V be a finite-dimensional vector space over F, and let $T : V \to V$ be a linear operator. Choose any vector $y \in V$. Then, keeping T and y fixed, we can define a map $f : V \to F$ by $f(x) = \langle Tx, y \rangle \; \forall x \in V$.

E E3) Show that f is a linear functional, i.e., $f \in V^*$.

By E3 and Theorem 2, $\exists$ a unique element $z \in V$ such that $f = \langle\ , z\rangle$, that is, $f(x) = \langle x, z\rangle \ \forall\ x \in V$, that is, $\langle Tx, y\rangle = \langle x, z\rangle \ \forall\ x \in V$.

Note that the choice of this vector z depends upon the fixed vector y. This is because if the fixed vector y is replaced by another vector y_1, we shall get another linear functional f_1 and f_1 will be represented as an inner product with some other vector z_1. Of course, you can see that f depends on T also!

So, for each $y \in V$, $\exists$ a unique vector $z \in V$, that depends only upon y, if we keep T fixed. Therefore, we get a function

$T^* : V \to V : T^*(y) = z$.

Then, we can write

$\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in V$ (since both are equal to $\langle x, z\rangle$).

We will look at some characteristics of the map T^* in the following two theorems. Henceforth, unless otherwise mentioned, **we will only deal with finite-dimensional inner product spaces.**

Theorem 3: If $(V, \langle\ ,\ \rangle)$ is an inner product space over the field **F** and $T \in A(V)$, then T^* is a linear transformation, i.e., $T^* \in A(V)$.

Proof: Choose $y_1, y_2 \in V$. Then, for any $x \in V$,

$$\begin{aligned}\langle x, T^*(y_1 + y_2)\rangle &= \langle Tx, y_1 + y_2\rangle, \text{ by definition.}\\ &= \langle Tx, y_1\rangle + \langle Tx, y_2\rangle\\ &= \langle x, T^*y_1\rangle + \langle x, T^*y_2\rangle, \text{ by definition.}\\ &= \langle x, T^*y_1 + T^*y_2\rangle\end{aligned}$$

This is true for any $x \in V$.

Therefore, $T^*(y_1 + y_2) = T^*(y_1) + T^*(y_2) \ \forall\ y_1, y_2 \in V$, by Unit 12 (Theorem 1).

Again, choose $y \in V$. Then, for any $x \in V$, and $\alpha \in F$,

$$\begin{aligned}\langle x, T^*(\alpha y)\rangle &= \langle Tx, \alpha y\rangle = \bar{\alpha}\langle Tx, y\rangle\\ &= \bar{\alpha}\langle x, T^*y\rangle\\ &= \langle x, \alpha T^*y\rangle,\end{aligned}$$

which implies that $T^*(\alpha y) = \alpha T^*(y)$.

Thus, we have shown that T^* is linear.

So, we have shown that given $T \in A(V)\ \exists\ T^* \in A(V)$, such that $\langle Tx, y\rangle = \langle x, T^*y\rangle$ for $x, y \in V$. Now, we will show that T^* is unique.

Theorem 4: If $(V, \langle\ ,\ \rangle)$ is an inner product space over F and $T \in A(V)$, then $\exists$ a unique $T^* \in A(V)$ for which

$\langle Tx, y\rangle = \langle x, T^*y\rangle$ for all $x, y \in V$.

Proof: Suppose T^* is not unique. Then there will exist at least two operators T_1^*, $T_2^* \in A(V)$ such that

$\langle Tx, y\rangle = \langle x, T_1^*y\rangle$

and $\langle Tx, y\rangle = \langle x, T_2^*y\rangle$

for all $x, y \in V$. This will mean that $\forall\ x, y \in V$

$\langle x, T_1^*y\rangle = \langle x, T_2^*y\rangle$. $\therefore$ $\langle x, T_1^*(y) - T_2^*(y)\rangle = 0 \ \forall\ x \in V$.

$\therefore$ $T_1^*y = T_2^*y$ for all $y \in V$.

This shows that $T_1^* = T_2^*$.

Theorem 4 allows us to give the following definition.

Definition: If $(V, \langle\ ,\ \rangle)$ is an inner product space over the field **F** and $T \in A(V)$, then the unique operator $T^* \in A(V)$ for which $\langle Tx, y\rangle = \langle x, T^*y\rangle$ holds for all $x, y \in V$, is called the **adjoint** of the operator **T**. (We also call T^* the adjoint operator.)

Let us look at some examples.

Example 2: Let $P_n(\mathbb{C})$ denote the vector space of all polynomials of degree ≤ n with complex coefficients. Show that we can define an inner product on $P_n(\mathbb{C}) = P_n$ as follows:
$\langle f, g\rangle = \sum_{i=0}^{n} a_i \overline{b_i}$ where $f = a_0 + a_1t + \ldots a_nt^n$ and $g = b_0 + b_1t + \ldots + b_nt^n$. Find T^* for the operator T defined by $(Tf)(t) = af(t)$, $a \in \mathbb{C}$.

Solution: Take $B = \{1, t, t^2, \ldots, t^n\}$ in Example 3 of Unit 12. Then you can see that $\langle\, ,\, \rangle$, defined above, is an inner product. Now for $f, g \in P_n$,

$\langle Tf, g\rangle = \langle af, g\rangle = a\langle f, g\rangle = \langle f, \bar{a}g\rangle$.

$\therefore$ $\langle f, T^*g\rangle = \langle f, \bar{a}g\rangle \forall f, g \in P_n$. $\therefore T^*g = \bar{a}g \forall g \in P_n$.

$\therefore$ we get $T^* : P_n \to P_n : T^*(f) = \bar{a}f$.

Example 3: Find D^* for the differential operator D, defined on P_n by $Df(t) = f'(t)$.

Solution: For $f = a_0 + a_1t + \ldots + a_nt^n$ and $g = b_0 + b_1t + \ldots + b_nt^n$, we have

$$\begin{aligned}\langle Df, g\rangle = \langle f', g\rangle &= \langle a_1 + 2a_2t + \ldots + na_nt^{n-1}, g\rangle \\ &= a_1\bar{b}_0 + 2a_2\bar{b}_1 + \ldots + na_n\bar{b}_{n-1} \\ &= \langle a_0 + a_1t + \ldots a_nt^n, b_0t + 2b_1t^2 + \ldots + nb_{n-1}t^n\rangle\end{aligned}$$

$$\begin{aligned}\therefore D^*(b_0 + b_1t + \ldots + b_nt^n) &= b_0t + 2b_1t^2 + \ldots + nb_{n-1}t^n \\ &= t(b_0 + 2b_1t + \ldots + nb_{n-1}t^{n-1})\end{aligned}$$

Try the following exercise now.

E E4) Obtain the adjoint of the operator

$T : \mathbf{R}^n \to \mathbf{R}^n : T(x_1, \ldots, x_n) = (x_1, 0, \ldots, 0)$.

Let us now look at some basic properties of the adjoint operator.

Theorem 5: Let $(V, \langle\, ,\, \rangle)$ be an inner product space over F. Then, for $S, T \in A(V)$, the following relations hold:

a) $I^* = I$, I being the identity operator.

b) $(S + T)^* = S^* + T^*$

c) $(\alpha T)^* = \bar{\alpha}T^*$, for any $\alpha \in F$.

d) $\langle T^*(y), x\rangle = \langle y, T(x)\rangle$, for all $x, y \in V$.

e) $T^{**} = T$ (T^{**} means $(T^*)^*$)

f) $T^*T = \mathbf{0}$ iff $T = \mathbf{0}$.

g) $(T \circ S)^* = S^* \circ T^*$.

Proof: We will prove (e), (f) and (g) here, assuming (a) to (d). We leave the proof of (a) – (d) to you (see E5).

e) Choose any two vectors $x, y \in V$. Then,

$\langle T^{**}(x), y\rangle = \langle (T^*)^*(x), y\rangle = \langle x, T^*(y)\rangle$, by (d).

$= \langle T(x), y\rangle$, by definition.

This is true for any $y \in V$.

$\therefore$ $T^{**}(x) = T(x) \forall x \in V$. Hence, $T^{**} = T$.

f) If $T^*T = \mathbf{0}$, then, for each $x \in V$, $T^*T(x) = \mathbf{0}$.

Hence, $\langle T^*T(x), y\rangle = 0$ for any $y \in V$.

Thus, for $y = x$ we get $0 = \langle T^*T(x), x\rangle = \langle T^*(T(x)), x\rangle$

$= \langle T(x), T(x)\rangle$, by (d)

$\Rightarrow T(x) = \mathbf{0}$, by IP2 (Unit 12).

Therefore, $T(x) = \mathbf{0}$ for each $x \in V$. Hence $T = \mathbf{0}$

Conversely, if $T = 0$ then $T(x) = 0 \ \forall x \in V$

$\Rightarrow T^*T(x) = 0 \ \forall x \in V$

$\Rightarrow T^*T = 0.$

g) For any $x, y \in V$, $\langle (T \circ S)^*(x), y\rangle = \langle x, (T \circ S)(y)\rangle$, by (d)

$= \langle x, T(S(y))\rangle$

$= \langle T^*(x), S(y)\rangle$, by (d).

$\langle S^*(T^*(x)), y\rangle$, by (d).

$= \langle (S^* \circ T^*)(x), y\rangle$

$\therefore (T \circ S)^*(x) = (S^* \circ T^*)(x)$ for any $x \in V$.

Hence, $(T \circ S)^* = S^* \circ T^*$.

To complete the proof of this theorem, try E5.

E E5) Prove (a) – (d) of Theorem 5.

$TT^* = 0 \Leftrightarrow T^* = 0$

Now, look closely at (e) and (f) of Theorem 5. They tell us that for any $T \in A(V)$

$TT^* = 0 \Leftrightarrow T^{**}T^* = 0$, since $T^{**} = T$.

$\Leftrightarrow T^* = 0$, by (f) applied to T^*.

Try the following exercises now.

E E6) Show that if $T = 0$, then so is T^*.

E E7) Show that the map $\phi: A(V) \to A(V) : \phi(T) = T^*$ is sesquilinear, that is, $\phi(S + T) = \phi(S) + \phi(T)$, and $\phi(\alpha S) = \bar{\alpha}\phi(S) \ \forall S, T \in A(V)$ and $\alpha \in F$.

E E8) Using Theorem 5, prove that if $T \in A(V)$ and T^{-1} exists then $(T^{-1})^* = (T^*)^{-1}$.

Now that you are familiar with the adjoint operator, let us look at some operators whose adjoints have special properties.

13.4 SOME SPECIAL OPERATORS

In this section we will define two types of transformations. They are classified according to the way their adjoints behave. The two types are self-adjoint operators and unitary operators.

13.4.1 Self-adjoint Operators

As the name indicates, the members of this class will consist of operators that are the same as their adjoints. We make a formal definition.

Definition: Let $(V, \langle , \rangle)$ be an inner product space over F and $T \in A(V)$. T is said to be **self-adjoint (or Hermitian)** if $T = T^*$.

Thus, if T is self-adjoint, then

$\langle Tx, y\rangle = \langle x, Ty\rangle = \overline{\langle Ty, x\rangle}$ for any $x, y \in V$.

If V is a real inner product space and T is self-adjoint, then the above condition is reduced to

$\langle Tx, y\rangle = \langle Ty, x\rangle$ (since $z = \bar{z} \ \forall \ z \in \mathbf{R}$).

In this case T is said to be **symmetric**.

Can you think of an example of a self-adjoint operator? Theorem 5 tells us that the identity operator is self-adjoint.

The following exercises deal with self-adjoint operators.

E E9) Define a function $f : \mathbf{R}^2 \to \mathbf{R}^2 : f(x, y) = (y, x)$. Show that f is self-adjoint.

E E10) If $S, T \in A(V)$ are self-adjoint, then show that $S \circ T$ is self-adjoint iff $S \circ T = T \circ S$, i.e., S and T commute. (Use Theorem 5.)

In Unit 10 you studied about the eigenvalues and eigenvectors of operators. Let us see what they look like in the case of self-adjoint operators.

Theorem 6: Let $(V, \langle , \rangle)$ be an inner product space and $T \in A(V)$ be self-adjoint. Then the eigenvalues of T are all real.

Proof: Let α be an eigenvalue of T. Then $\exists\, v \in V, v \neq 0$, such that $T(v) = \alpha v$. We want to show that $\alpha \in \mathbf{R}$. Now,

$$\begin{aligned}\alpha\langle v, v\rangle &= \langle \alpha v, v\rangle = \langle Tv, v\rangle \\ &= \langle v, T^* v\rangle = \langle v, Tv\rangle, \text{ since } T = T^*, \\ &= \langle v, \alpha v\rangle = \bar{\alpha}\langle v, v\rangle.\end{aligned}$$

Since $\langle v, v\rangle \neq 0$, we get $\bar{\alpha} = \alpha$. This means that $\alpha \in \mathbf{R}$.

The following exercise tells us something about **skew-Hermitian** operators.

$T \in A(V)$ is called **skew-Hermitian** if $T^* = -T$

E E11) Let V be a complex inner product space and $T \in A(V)$ such that $T^* = -T$. Show that

a) iT is self-adjoint, where $i = \sqrt{-1}$.

b) the eigenvalues of T are purely imaginary numbers or 0.

c) eigenvectors of T corresponding to distinct eigenvalues are mutually orthogonal.

We will now prove a useful result about self-adjoint operators.

Theorem 7: Let $(V, \langle\ ,\ \rangle)$ be an inner product space and $T \in A(V)$ be self-adjoint. Then $T = 0$ iff $\langle Tx, x\rangle = 0 \ \forall x \in V$.

Proof: For any operator T.

$T = 0 \Rightarrow Tx = 0 \ \forall x \in V \Rightarrow \langle Tx, x\rangle = 0 \ \forall x \in V$.

Conversely, assume that $\langle Tx, x\rangle = 0 \ \forall x \in V$.

Then $\langle T(x+y), x+y\rangle = 0 \ \forall x, y \in V$.

$\Rightarrow \langle Tx, y\rangle + \langle Ty, x\rangle = 0 \ \forall x, y \in V$. ......(1)

$\Rightarrow \langle Tx, y\rangle + \langle y, Tx\rangle = 0 \ \forall x, y \in V$, $\because T = T^*$.

$\Rightarrow \langle Tx, y\rangle + \overline{\langle Tx, y\rangle} = 0 \ \forall x, y \in V$.

$\Rightarrow \text{Re}\,\langle Tx, y\rangle = 0 \ \forall x, y \in V$.

Now 2 cases arise — $F = \mathbf{R}$ or $F = \mathbf{C}$.

If $F = \mathbf{R}$, then $\langle Tx, y\rangle = \text{Re}\,\langle Tx, y\rangle = 0 \ \forall x, y \in V$.

$\therefore T = 0$.

If $F = \mathbf{C}$, then $\langle T(ix + y), ix + y\rangle = 0 \ \forall x, y \in V$ gives us

$\langle Tx, y\rangle - \langle Ty, x\rangle = 0 \ \forall x, y \in V$.

This, with (1), gives us $\langle Tx, y\rangle = 0 \ \forall x, y \in V$

$\therefore$, again, $T = 0$.

This theorem will come in useful in the next sub section, where we look at another type of linear transformation.

13.4.2 Unitary Operators

We will now study the class of operators which satisfy the condition $T^* = T^{-1}$. First, a definition.

Definition: If $(V, \langle\, , \rangle)$ is an inner product space over **F** and $T \in A(V)$, then T is called unitary if

$TT^* = I = T^*T$.

Thus, T is unitary if and only if $T^* = T^{-1}$.

If **F** = **R**, a unitary operator is also called **orthogonal**.

Can you think of an example of a unitary operator? Does the identity operator satisfy the equation $II^* = I = I^*I$? Yes.

Another example is $f : \mathbf{R}^2 \to \mathbf{R}^2 : f(x, y) = (y, x)$.

From E9 you know that $f = f^*$. Also

$ff^*(x_1, x_2) = f(x_2, x_1) = f(x_1, x_2) \quad \therefore ff^* = I$.

Similarly $f^*f = I$. $\therefore$ f is unitary.

In both these examples you may have noticed that the operators are also self-adjoint. The following exercise will give you an example of a unitary operator which is not self-adjoint.

E E12) Show that the operator

$T : \mathbf{R}^3 \to \mathbf{R}^3 : T(x_1, x_2, x_3) = (x_3, x_1, x_2)$

is not self-adjoint, but it is unitary.

(Hint: Show that $T^* = T^2$ and $T^3 = I$.)

We will now prove a theorem that shows the utility of a unitary (orthogonal) operator.

Theorem 8: If $(V, \langle\, , \rangle)$ is an inner product space over F and $T \in A(V)$, then the following conditions are equivalent.

a) $T^*T = I$.

b) $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in V$.

c) $\|Tx\| = \|x\|$ for all $x \in V$.

Proof: We shall prove (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (a). This will show that all three statements are equivalent.

(a) $\Rightarrow$ (b): Assume (a). Then, for any $x, y \in V$, $\langle x, y\rangle = \langle Ix, y\rangle$

$= \langle T^*Tx, y\rangle = \langle Tx, Ty\rangle$.

Thus (b) holds.

(b) $\Rightarrow$ (c): If (b) holds for all x, $y \in V$, then it also holds when x = y. This means that, $\forall x \in V$,

$\langle Tx, Tx\rangle = \langle x, x\rangle$ or $\|Tx\|^2 = \|x\|^2$.

$\therefore \|Tx\| = \|x\| \forall x \in V$. Thus, (c) holds.

(c) $\Rightarrow$ (a): If (c) holds, then

$\langle Tx, Tx\rangle = \langle x, x\rangle$ for all $x \in V$.

$\Rightarrow \langle T^*T_x, x\rangle = \langle x, x\rangle$ for all $x \in V$.

$\Rightarrow \langle T^*T_x, x\rangle - \langle x, x\rangle = 0$ for all $x \in V$.

$\Rightarrow \langle (T^*T - I)x, x\rangle = 0$ for all $x \in V$.

$\Rightarrow T^*T - I = 0$ (by Theorem 7, since T^*T-I is self-adjoint)

$\Rightarrow T^*T = I$, which shows that (a) holds.

Note: Theorem 8 says that T is a unitary operator iff

i) $\langle Tx, Ty\rangle = \langle x, y\rangle \ \forall x, y \in V$, that is, **T preserves inner products.**

ii) $\|Tx\| = \|x\| \ \forall x \in V$, that is, **T preserves the length of a vector.**

You will learn about some properties of unitary operators from the following exercises.

E E13) If V is a given inner product space over C and $S, T \in A(V)$ are unitary operators, show that

a) $S \circ T$ is a unitary operator.

b) αT is a unitary operator for $\alpha \in C$ iff $|\alpha| = 1$.

E E14) Show that the characteristic roots of a unitary operator have absolute value 1.

Let us now talk about the action of a unitary operator on an orthonormal basis. From Unit 12 (Theorem 7) you know that $(V, \langle\, ,\, \rangle)$ has an orthonormal basis. The following theorem characterises unitary operators in terms of their action on an orthonormal basis.

Theorem 9: Let $(V, \langle\, ,\, \rangle)$ be an inner product space over F of dimension n. Then $T \in A(V)$ is unitary if and only if T maps an orthonormal basis of V onto an orthonormal basis of V.

Proof: Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Then $\langle e_i, e_j\rangle = 0$ if $i \neq j$ and $\langle e_i, e_i\rangle = 1 \ \forall i, j = 1, \ldots, n$.

We will first show that if T is unitary then $\{Te_1, \ldots, Te_n\}$ is an orthonormal basis of V. Now, since T preserves inner products, we get $\langle Te_i, Te_j\rangle = 0$ for $i \neq j$, and $\langle Te_i, Te_i\rangle = 1 \ \forall i, j = 1, \ldots, n$. Also, since T is invertible (in fact, $T^{-1} = T^*$), you know from Unit 5 that T maps a basis to a basis. Hence, $\{Te_1, \ldots, Te_n\}$ is an orthonormal basis.

Conversely, we will show that if $B = \{Te_1, \ldots, Te_n\}$ is an orthonormal basis then T is unitary. For this, consider

$$x = \sum_{i=1}^{n} \alpha_i e_i, \quad y = \sum_{i=1}^{n} \beta_i e_i \quad \text{in } V, \text{ where } \alpha_i, \beta_i \in F \ \forall i = 1, \ldots, n.$$

Then

$$\langle Tx, Ty\rangle = \left\langle \sum_i \alpha_i T(e_i), \sum_j \beta_j T(e_j)\right\rangle$$

$$= \sum_i \sum_j \alpha_i \bar{\beta}_j \langle Te_i, Te_j\rangle$$

$$= \sum \alpha_i \bar{\beta}_i, \text{ since } B \text{ forms an orthonormal basis.}$$

Also, $\langle x, y\rangle = \sum_i \alpha_i \overline{\beta}_i$

Thus, $\langle Tx, Ty\rangle = \langle x, y\rangle \; \forall x, y \in V$.

Hence, by Theorem 8 we can say that T is unitary.

We will use Theorem 9 to solve the following example.

Example 4: Let $(V, \langle\, , \rangle)$ be a real inner product space of dimension 2. Obtain an orthogonal operator $T \in A(V)$ such that $\langle Tx, x\rangle = 0 \; \forall x \in V$.

Solution: Let $\{e_1, e_2\}$ be an orthonormal basis of V. Then, so is $\{e_2, -e_1\}$. If we define $T \in A(V)$ by $T(e_1) = e_2$ and $T(e_2) = -e_1$, by Theorem 9 we know that T is orthogonal. Also, $\langle Te_1, e_1\rangle = 0 = \langle Te_2, e_2\rangle$.

Now take any $x \in V$. Then $\exists\, a, b \in F$ such that $x = ae_1 + be_2$. What is $\langle Tx, x\rangle$? It is

$\langle T(ae_1 + be_2), ae_1 + be_2\rangle$

$= \langle ae_2 - be_1, ae_1 + be_2\rangle$

$= ab - ab = 0$. Thus, T is the required operator.

Note that this example shows us that **Theorem 7 is false if T is not self-adjoint.**

Try the following exercise now.

E E15) If $T \in A(V)$ be such that $T^2 = I$, show that T is Hermitian if and only if T is unitary.

So far we have been discussing various kinds of operators. You may have wondered about their matrix analogues. That is what we will discuss in the next section. But, before going further revise Unit 7.

13.5 HERMITIAN AND UNITARY MATRICES

In previous blocks you have seen the inter-relationship between operators and matrices representing them. In this section we will show you the link between self-adjoint operators and Hermitian matrices, and between unitary operators and unitary matrices.

13.5.1 Matrix of the Adjoint Operator

Let $(V, \langle\, , \rangle)$ be an inner product space over F. Given the matrix representation of an operator $T \in A(V)$, a natural problem that we can ask is: what is the matrix representation of its adjoint T^*?

To solve it let us consider an orthonormal basis $B = \{e_1, \ldots, e_n\}$ of V. Let $[T]_B = [a_{ij}]$ and $[T^*]_B = [b_{ij}]$. Then we know that

$T(e_i) = a_{1i}e_1 + a_{2i}e_2 + \ldots + a_{ni}e_n \; \forall i = 1, \ldots, n.$

and $T^*(e_i) = b_{1i}e_1 + \ldots + b_{ni}e_n \; \forall i = 1, \ldots, n.$

Now for $e_i, e_j \in B$, we have

$\langle Te_i, e_j\rangle = \langle e_i, T^*e_j\rangle$

$$\Rightarrow \left\langle \sum_{k=1}^{n} a_{ki}e_k, e_j \right\rangle = \left\langle e_i, \sum_{k=1}^{n} b_{kj}e_k \right\rangle$$

$$\Rightarrow \sum_{k=1}^{n} a_{ki}\langle e_k, e_j\rangle = \sum_{k=1}^{n} \overline{b_{kj}}\langle e_i, e_k\rangle$$

$$\Rightarrow a_{ji} = \overline{b_{ij}}, \text{ since } \langle e_i, e_j\rangle = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases}$$

Thus, we have proved the following result.

Theorem 10: Let V be an inner product space over F(F = **R** or **C**) of dimension n, and $T \in A(V)$ have the matrix representation $[a_{ij}]$ with respect to a given orthonormal basis B. Then the matrix representation of the adjoint T^* of T with respect to the same basis is the matrix $[b_{ij}]$, where $b_{ij} = \overline{a_{ji}}$.

Note: When **F** = **R**, then $b_{ij} = a_{ji}$.

Recall, from Unit 7, that given a matrix $A = [a_{ij}]$, its conjugate transpose is the matrix $A^* = [a_{ij}^*]$, where $a_{ij}^* = \overline{a_{ji}}$, i.e., $A^* = \overline{A^t}$.

Thus, Theorem 10 says that:

If $A = [a_{ij}]$ is the matrix representation of $T \in A(V)$ with respect to B, then the matrix representation of the adjoint T^* with respect to B is $A^* = \overline{A^t}$.

For example, if $D : P_2 \to P_2$ is the differential operator, then its matrix with respect to the orthonormal basis $B = \{1, x, x^2\}$ is

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \quad \therefore \; [D^*]_B = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}$$

Try the following exercises about the conjugate transposes of matrices.

E E16) Show that

a) $(AB)^* = B^*A^*$ for any two $n \times n$ matrices A and B.
(**Hint:** Show that the (i, j)th elements of $(AB)^*$ and B^*A^* are the same.)

b) If an $n \times n$ matrix A is invertible, then so is A^*; and $(A^*)^{-1} = (A^{-1})^*$.

Now let us look at the matrix of a self-adjoint operator.

13.5.2 Hermitian Matrix

Recall, from Unit 7 (Sec. 7.3.3), that a matrix A is said to be **Hermitian** if it is equal to its conjugate transpose, that is, if $A = A^*$. The following result tells us that the matrix of a Hermitian operator is Hermitian.

Theorem 11: Let V be an inner product space over **F** and $T \in A(V)$. Let the matrix representation of T with respect to an orthonormal basis $B = \{e_1, \ldots, e_n\}$ be A. Then T is self-adjoint iff A is Hermitian.

Proof: Let $[T]_B = A = [a_{ij}]$. Then, by Theorem 10,
$[T^*]_B = [b_{ij}]$, where $b_{ij} = \overline{a_{ji}}$. That is, $[T^*]_B = A^*$.

If T is self-adjoint, i.e., $T = T^*$. Therefore, $[T]_B = [T^*]_B$. Therefore, $A = A^*$, which means A is Hermitian.

Conversely, if A is Hermitian, then $A = A^*$. Therefore,

$a_{ij} = a^*_{ij} = \overline{a_{ji}} \ \forall i, j = 1, \ldots, n.$

Now, by definition, $T(e_i) = \sum_{j=1}^{n} a_{ji} e_j$. Therefore,

$$\langle Te_i, e_k \rangle = \left\langle \sum_{j=1}^{n} a_{ji} e_j, e_k \right\rangle = \sum_{j=1}^{n} a_{ji} \langle e_j, e_k \rangle = a_{ki} = \overline{a_{ik}} \ \forall i, k = 1, \ldots, n.$$

Also $T^*(e_i) = \sum_{j=1}^{n} a^*_{ji} e_j \ \forall i = 1, 2, \ldots, n.$

$\therefore \langle T^* e_i, e_k \rangle = a^*_{ki} = \overline{a_{ik}} \ \forall i, k = 1, \ldots, n.$

$\therefore \langle Te_i, e_k \rangle = \langle T^* e_i, e_k \rangle \ \forall i, k = 1, \ldots, n.$

This means that $T = T^*$, that is, T is self-adjoint.

Thus, the theorem is proved.

So, by Theorem 11 we know that the matrix of the operator in E9, with respect to the standard basis, is a Hermitian matrix. That is, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is Hermitian.

Theorem 11 also tells us that the following Hermitian matrices, treated as operators, are self-adjoint:

$[3]$, $\begin{bmatrix} 2 & 1+i \\ 1-i & 0 \end{bmatrix}$

and $\begin{bmatrix} k & a+ib & c+id \\ a-ib & m & e+if \\ c-id & e-if & n \end{bmatrix}$, where a, b, c, d, e, f, k, m, n $\in$ **R**.

You may like to try the following exercises now.

E E17) How many characteristic roots of a Hermitian matrix are purely imaginary? (**Hint:** Use Theorem 6.)

E E18) Show that a triangular Hermitian matrix is a diagonal matrix.

E E19) Show that the matrix A of a skew–Hermitian operator $T \in A(V)$ (i.e., $T = -T^*$), with respect to an orthonormal basis of V is skew–Hermitian (i.e., $A = -A^*$).

We will now introduce you to the matrix corresponding to a unitary operator.

13.5.3 Unitary (Orthogonal) Matrix

Remember, whenever we discuss unitary operators, we include orthogonal operators, that is, the case $F = R$. We will lead you to the definition of a unitary matrix, via the following theorem.

Theorem 12: Let V be an inner product space over **F** with dim V = n. Let $U \in A(V)$ have a matrix representation $A = [a_{ij}]$, with respect to an orthonormal basis B of V. If U is unitary, then

The Kronecker delta, δ_{ij}, is defined by $\delta_{ij} = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j. \end{cases}$

a) $\sum_{k=1}^{n} a_{ik}\overline{a_{jk}} = \delta_{ij}$

b) $\sum_{k=1}^{n} \overline{a_{ki}}a_{kj} = \delta_{ij}$

$\forall i, j = 1, \ldots, n.$

Proof: U has the matrix representation $A = [a_{ij}]$, with respect to B. Therefore, U^* has the matrix representation $A^* = [a_{ij}^*]$ with respect to B, $a_{ij}^* = \overline{a_{ji}}$. Since U is unitary, $UU^* = I = U^*U$. Therefore, $AA^* = I = A^*A$.
That is $[a_{ij}][a_{ij}^*] = [\delta_{ij}] = [a_{ij}^*][a_{ij}]$

Now, $[a_{ij}][a_{ij}^*] = [\delta_{ij}]$

$$\Rightarrow \sum_{k=1}^{n} a_{ik}a_{kj}^* = \delta_{ij}$$

$$\Rightarrow \sum_{k=1}^{n} a_{ik}\overline{a_{jk}} = \delta_{ij}$$

Similarly, $[a_{ij}^*][a_{ij}] = [\delta_{ij}]$.

$$\Rightarrow \sum_{k=1}^{n} \overline{a_{ki}}a_{kj} = \delta_{ij}.$$

The above result leads us to the following definition.

Definition: If A is a given $n \times n$ matrix with entries in a field **F**, then A is said to be a **unitary matrix** (an **orthogonal matrix**, if $F = R$) if $AA^* = I = A^*A$.

Thus, Theorem 12 says that:

The matrix representation of any unitary (or orthogonal) operator on an inner product space V, with respect to an orthonormal basis, is a unitary (or orthogonal) matrix.

Example 5: Show that the matrix $A = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}$ is not orthogonal.

Solution: $A^* = A^t$ in this case, since the entries of A are real.

Thus, $AA^* = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 13 \end{bmatrix} \neq I$

This means that A is not orthogonal.

The following exercises will give you some examples of unitary matrices.

E E20) Which of the following matrices are unitary?

$\begin{bmatrix} 0 & 1 & i \\ -1 & i & 2 \end{bmatrix}, \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix}.$

E21) Is $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$ orthogonal?

We will now derive a basic property of unitary (and orthogonal) matrices.

Theorem 13: For a square matrix A over **C** the following are equivalent.

a) A is unitary.

b) The rows of A form an orthonormal set of vectors.

c) The columns of A form an orthonormal set of vectors.

Proof: We will prove that (a) ⇔ (b) and (a) ⇔ (c).

(a) ⇔ (b): Let $A = [a_{ij}] = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix}$, where R_i is the ith row of A. Then $\overline{R_i^t}$ will be the ith column of A^*.

$\therefore AA^* = I \Leftrightarrow \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} \left[\overline{R_1^t}, \ldots, \overline{R_n^t}\right] = I.$

$$\Leftrightarrow \begin{bmatrix} R_1\overline{R_1^t} & R_1\overline{R_2^t} & \ldots\ldots & R_1\overline{R_n^t} \\ \vdots & \vdots & \ldots\ldots & \vdots \\ \vdots & \vdots & \ldots\ldots & \vdots \\ R_n\overline{R_1^t} & R_n\overline{R_2^t} & \ldots\ldots & R_n\overline{R_n^t} \end{bmatrix} = I.$$

$\Leftrightarrow R_i\overline{R_j^t} = \delta_{ij} \ \forall\ i, j = 1, \ldots, n.$

$$\Leftrightarrow (a_{i1}, \ldots, a_{in}) \begin{bmatrix} \overline{a_{j1}} \\ \vdots \\ \overline{a_{jn}} \end{bmatrix} = \delta_{ij}$$

$$\Leftrightarrow \sum_{k=1}^{n} a_{ik}\overline{a_{jk}} = \delta_{ij}.$$

$\Leftrightarrow$ the set of vectors $\{(a_{i1}, \ldots, a_{in}) \mid i = 1, \ldots, n\}$ are orthonormal.

$\Leftrightarrow$ the rows of A are orthonormal.

Hence, we have proved that (a) $\Leftrightarrow$ (b).

Similarly, using the fact that $A^*A = I$, we can prove that (a) $\Leftrightarrow$ (c). Hence, we have proved the theorem.

Note: Just as we have proved Theorem 13, we can prove that a real square matrix is orthogonal iff its rows (or columns) form an orthonormal set of vectors.

You can apply what we have just said to solve the following exercise.

E E22) Consider the matrix representing the linear operator $T : \mathbf{R}^3 \to \mathbf{R}^3 : T(x, y, z) = (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta, z)$, with respect to the standard basis. Do its columns form an orthonormal set of vectors?

Now let us look at **real matrices only** for the rest of the section.

Recall, from Unit 7, that a matrix A is symmetric if $A = A^t$. In Unit 10 you also came across the concept of similar matrices. We now define an allied concept.

Definition: Two square matrices A and B, of the same order, are said to be **orthogonally similar** if $A = P^{-1}BP$, for some orthogonal matrix P.

Remember that if P is orthogonal, then it is invertible, and its inverse is P^t. Thus, **A and B are orthogonally similar if $A = P^tBP$, for an orthogonal matrix P.**

Let us consider an example.

Example 6: Show that $\begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}$ and $\begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix}$ are orthogonally similar.

Solution: Suppose $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is an orthogonal matrix satisfying

$$\begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} = P^t \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} P.$$

Then we have $\begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$= \begin{bmatrix} -2a + 2c & -a + c \\ -2b + 2d & -b + d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} (2a+c)(c-a) & (2b+d)(c-a) \\ (2a+c)(d-b) & (2b+d)(d-b) \end{bmatrix}$$

Solving the equations

$$1 = (2a+c)(c-a)$$
$$-1 = (2a+c)(d-b)$$
$$2 = (2b+d)(c-a)$$
$$-2 = (2b+d)(d-b),$$

we get

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

$$\therefore \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^t \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\text{or } \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}^t \begin{bmatrix} -2 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}.$$

Check that these equalities do hold, by multiplying the right hand side.

This example shows that there can be several orthogonal matrices P such that $A = P^tBP$.

Now we shall use an orthogonal matrix to diagonalise a real symmetric matrix. In Unit 10 you have studied about diagonalising matrices. Theorem 5 of Unit 10 gives you a practical method of diagonalising a square matrix. We will use this theorem to prove the following result.

Theorem 14: Let A be a real symmetric matrix of order n with distinct eigenvalues $\alpha_1, \ldots, \alpha_n$. Let $X_1, \ldots, X_n \in V_n(R)$ be normalised eigenvectors (see Sec. 12.3) of A corresponding to $\alpha_1, \ldots, \alpha_n$, respectively. Let $P = (X_1, \ldots, X_n)$. Then

a) P is orthogonal.

b) P^tAP is the diagonal matrix, diag $(\alpha_1, \ldots, \alpha_n)$.

Proof: a) We will first show that $\{X_1, \ldots, X_n\}$ is an orthonormal set in $V_n(R)$. Remember that the standard inner product in $V_n(R)$ is given by $X \cdot Y = \sum_{i=1}^{n} x_i y_i = X^tY$

$$\forall X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \text{ in } V_n(R).$$

$$\begin{aligned}
\text{Now, } (\alpha_1 - \alpha_2)(X_1 \cdot X_2) &= \alpha_1(X_1 \cdot X_2) - \alpha_2(X_1 \cdot X_2) \\
&= (\alpha_1 X_1 \cdot X_2) - (X_1 \cdot \alpha_2 X_2) = (AX_1 \cdot X_2) - (X_1 \cdot AX_2) \\
&= (AX_1)^t X_2 - X_1^t AX_2 \\
&= X_1^t AX_2 - X_1^t AX_2 \quad (\text{since } A^t = A) \\
&= 0
\end{aligned}$$

Since $\alpha_1 \neq \alpha_2$, we get $X_1 \cdot X_2 = 0$.

Similarly $X_i \cdot X_j = 0 \ \forall i \neq j$.

Also $\|X_i\| = 1 \ \forall i = 1, \ldots, n$, since the X_i's are normalised vectors.

Therefore, $\{X_1, X_2, \ldots, X_n\}$ is an orthonormal set.

Therefore, by Theorem 13, P is orthogonal.

b) From Unit 10 (Theorem 5) you know that $P^{-1}AP = \text{diag}(\alpha_1, \ldots, \alpha_n)$. That is, $P^tAP =$ diag $(\alpha_1, \ldots, \alpha_n)$.

What Theorem 14 says is that **any real symmetric n × n matrix with n distinct eigenvalues is orthogonally similar to a diagonal matrix**. This theorem has important geometrical applications in the study of quadrics. You will see the connection in Unit 15.

Note: Though we have proved Theorem 14 for real symmetric matrices with distinct eigenvalues, it is true for any real symmetric matrix. That is, **any real symmetric matrix is orthogonally similar to a diagonal matrix**. The proof of this result is beyond the scope of this course.

Let us consider an example of how to use Theorem 14.

Example 7: Reduce $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$ to diagonal form.

Solution: The matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & -1 \end{bmatrix}$ is a real symmetric matrix. Its characteristic equation is $\begin{vmatrix} \lambda - 1 & -1 & -1 \\ -1 & \lambda - 1 & 1 \\ -1 & 1 & \lambda + 1 \end{vmatrix} = 0$. This shows us that the eigenvalues of A are 1, 2, −2.

Eigenvectors corresponding to them are (1, −1, 1), (1, 1, 0) and (−1, 1, 2), respectively. Therefore, the normalised eigenvectors are $(1/\sqrt{3})(1, -1, 1)$, $(1/\sqrt{2})(1, 1, 0)$, $(1/\sqrt{6})(-1, 1, 2)$. These vectors give us the orthogonal matrix

$$P = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ -1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}$$

Then, we get $P^tAP = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$.

Do try the following exercise now.

E E23) Reduce $\begin{bmatrix} 7 & -1 & -10 \\ -1 & 7 & 10 \\ -10 & 10 & -2 \end{bmatrix}$ to diagonal form.

(Its eigenvalues are 6, −12 and 18.)

Let us end with summarising what we have covered in this unit.

13.6 SUMMARY

As in the previous unit, the vector spaces considered in this unit are all defined over the fields C or R. We made the following points in this unit.

1) Any linear functional on an inner product space is represented by the inner product with a fixed vector.
2) The definition and properties of the adjoint of an operator defined on an inner product space.
3) The definition and properties of a self-adjoint operator.
4) The definition and properties of a unitary (orthogonal) operator.
5) A self-adjoint operator on an inner product space is represented by a Hermitian matrix, with respect to an orthonormal basis of the underlying space.

6) A unitary (orthogonal) transformation on an inner product space is represented by a unitary (orthogonal) matrix, with respect to an orthonormal basis of the underlying space.
7) A matrix is unitary (orthogonal) iff its rows form an orthonormal set of vectors iff its columns form an orthonormal set of vectors.
8) Any real symmetric matrix is orthogonally similar to a diagonal matrix.

13.7 SOLUTIONS/ANSWERS

E1) For any $x_1, x_2 \in \mathbf{R}^2$, we have

$f_y(x_1 + x_2) = \langle x_1 + x_2, y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle$

$= f_y(x_1) + f_y(x_2)$.

Also, for any $a \in \mathbf{R}$ and $x \in \mathbf{R}^2$,

$f_y(ax) = \langle ax, y\rangle = af_y(x)$.

$\therefore f_y$ is a linear functional on $\mathbf{R}^2$.

E2) $\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ is an orthonormal basis of $\mathbf{C}^3$.

Now $f(1, 0, 0) = \frac{1}{3} = f(0, 1, 0) = f(0, 0, 1)$.

$\therefore$, as in the proof of Theorem 2,

$y = \frac{1}{3}(1, 0, 0) + \frac{1}{3}(0, 1, 0) + \frac{1}{3}(0, 0, 1) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ is what we want. To check whether y is the required vector you must ensure that $f(z) = \langle z, y\rangle \ \forall z \in \mathbf{C}^3$.

E3) For x_1 and $x_2 \in V$, we have

$f(x_1 + x_2) = \langle T(x_1 + x_2), y\rangle = \langle Tx_1 + Tx_2, y\rangle = \langle Tx_1, y\rangle + \langle Tx_2, y\rangle$

$= f(x_1) + f(x_2)$

Also, for $a \in \mathbf{F}$ and $x \in V$,

$f(ax) = \langle T(ax), y\rangle = \langle aTx, y\rangle = a\langle Tx, y\rangle = af(x)$.

$\therefore f \in V^*$.

E4) $\langle T(x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle = \langle (x_1, 0, \ldots, 0), (y_1, \ldots, y_n)\rangle$

$= x_1 y_1 = \langle (x_1, \ldots, x_n), (y_1, 0, \ldots, 0)\rangle$.

$\therefore T^*(y_1, \ldots, y_n) = (y_1, 0, \ldots, 0)$

$\therefore T^* = T$ in this case.

E5) a) For any $x, y \in V$, $\langle I(x), y\rangle = \langle x, y\rangle$

$\Rightarrow \langle x, I^*y\rangle = \langle x, y\rangle$

$\Rightarrow I^*(y) = y \ \forall y \in V \Rightarrow I^* = I$.

b) For any $x, y \in V$,

$\langle (S + T)(x), y\rangle = \langle Sx + Tx, y\rangle = \langle Sx, y\rangle + \langle Tx, y\rangle$

$= \langle x, S^*y\rangle + \langle x, T^*y\rangle$

$= \langle x, S^*y + T^*y\rangle$

$= \langle x, (S^* + T^*)y\rangle$.

$\therefore (S + T)^* = S^* + T^*$.

c) For any $x, y \in V$,

$\langle (\alpha T)(x), y\rangle = \langle \alpha Tx, y\rangle = \alpha\langle Tx, y\rangle$

$= \alpha\langle x, T^*y\rangle$

$= \langle x, (\bar{\alpha}T^*)(y)\rangle$

$\therefore (\alpha T)^* = \bar{\alpha}T^*$.

d) $\langle T^*y, x\rangle = \overline{\langle x, T^*y\rangle} = \overline{\langle Tx, y\rangle} = \langle y, Tx\rangle$.

E6) $T = 0 \Rightarrow Tx = 0 \ \forall x \in V \Rightarrow \langle Tx, y\rangle = 0 \ \forall x, y \in V$.

$\Rightarrow \langle x, T^*y\rangle = 0 \ \forall x, y \in V$.

In particular, for $x = T^*(y) \in V$, we get

$\langle T^* y, T^* y\rangle = 0 \ \forall\ y \in V.$

$\Rightarrow T^* y = 0 \ \forall\ y \in V \Rightarrow T^* = 0.$

E7) By Theorem 5(b), $\phi(S+T) = \phi(S) + \phi(T)$.

By Theorem 5(c), $\phi(\alpha S) = \bar{\alpha}\phi(S)$.

E8) By Theorem 5, $(T \cdot T^{-1})^* = (T^{-1})^* \cdot (T^*)$

$\Rightarrow I^* = (T^{-1})^* \cdot (T^*)$

$\Rightarrow I = (T^{-1})^* \cdot (T^*)$

Similarly, $T^*(T^{-1})^* = I$. $\therefore (T^{-1})^* = (T^*)^{-1}$.

E9) Now $\langle f(x_1, x_2), (y_1, y_2)\rangle = \langle (x_2, x_1), (y_1, y_2)\rangle$

$= x_2 y_1 + x_1 y_2$

$= x_1 y_2 + x_2 y_1$

$= \langle (x_1, x_2), (y_2, y_1)\rangle$

$\therefore f^*(y_1, y_2) = (y_2, y_1) = f(y_1, y_2) \ \forall\ (y_1, y_2) \in \mathbf{R}^2.$

$\therefore f^* = f.$

E10) $S \circ T = (S \circ T)^*$

$\Leftrightarrow S \circ T = T^* \circ S^* = T \circ S$, since $S = S^*$ and $T = T^*$.

E11) a) $(iT)^* = \bar{i}T^* = (-i)(-T) = iT$.

b) Let $\alpha \in \mathbf{C}$ be an eigenvalue of T. Then $\exists\ 0 \neq v \in V$ such that $Tv = \alpha v$. We will show that $\bar{\alpha} = -\alpha$.

Now $\alpha\langle v, v\rangle = \langle \alpha v, v\rangle = \langle Tv, v\rangle = \langle v, T^* v\rangle$

$= \langle v, -Tv\rangle = -\langle v, \alpha v\rangle$

$= -\bar{\alpha}\langle v, v\rangle.$

$\therefore \alpha = -\bar{\alpha} \Rightarrow \alpha = 0$ or α is purely imaginary.

c) Let $\alpha, \beta \in \mathbf{C}$ be distinct eigenvalues of T. Let $v, w \in V$ be eigenvectors corresponding to α and β, respectively. Then $Tv = \alpha v$ and $Tw = \beta w$.

Now $\alpha\langle v, w\rangle = \langle \alpha v, w\rangle = \langle Tv, w\rangle = \langle v, T^* w\rangle$

$= -\langle v, Tw\rangle = -\langle v, \beta w\rangle = -\bar{\beta}\langle v, w\rangle$

$= \beta\langle v, w\rangle$ ($\because \bar{\beta} = -\beta$ from (b) above).

$\Rightarrow (\alpha - \beta)\langle v, w\rangle = 0 \Rightarrow \langle v, w\rangle = 0$, since $\alpha \neq \beta$.

$\Rightarrow$ v is orthogonal to w.

E12) $\langle T(x_1, x_2, x_3), (y_1, y_2, y_3)\rangle = \langle (x_3, x_1, x_2), (y_1, y_2, y_3)\rangle$

$= x_3 y_1 + x_1 y_2 + x_2 y_3 = x_1 y_2 + x_2 y_3 + x_3 y_1$

$= \langle (x_1, x_2, x_3), (y_2, y_3, y_1)\rangle$

$\therefore T^*(y_1, y_2, y_3) = (y_2, y_3, y_1) = T^2(y_1, y_2, y_3) \ \forall\ (y_1, y_2, y_3) \in \mathbf{R}^3.$

$\therefore T^* = T^2 \neq T.$

Also $T^3(x) = x \ \forall\ x \in \mathbf{R}^3$. $\therefore T^2 = T^{-1}$, i.e., $T^* = T^{-1}$.

$\therefore$ T is unitary.

E13) a) $(ST)^* = T^* S^* = T^{-1} S^{-1} = (ST)^{-1}$.

b) $(\alpha T)^* = (\alpha T)^{-1} \Leftrightarrow \bar{\alpha}T^* = \alpha^{-1}T^{-1} \Leftrightarrow \alpha^{-1}T^{-1} = \alpha^{-1}T^{-1}$.

$\Leftrightarrow \bar{\alpha} = \alpha^{-1} \Leftrightarrow \alpha\bar{\alpha} = 1 \Leftrightarrow |\alpha| = 1.$

E14) Let α be a characteristic root, i.e., an eigenvalue of a unitary operator $T \in A(V)$. Then $\exists\ 0 \neq v \in V$ such that $T(v) = \alpha v$.

Now $\alpha\langle v, v\rangle = \langle \alpha v, v\rangle = \langle Tv, v\rangle = \langle v, T^* v\rangle = \langle v, T^{-1} v\rangle$

$= \langle v, \alpha^{-1} v\rangle$ ($\because T^{-1} v = \alpha^{-1} v$).

$= \overline{\alpha^{-1}}\langle v, v\rangle.$

$\therefore \alpha = \overline{\alpha^{-1}} \Rightarrow \alpha\bar{\alpha} = 1 \Rightarrow |\alpha| = 1.$

E15) $T^2 = I \Leftrightarrow T = T^{-1}$, Now,

T is Hermitian $\Leftrightarrow T = T^* \Leftrightarrow T^{-1} = T^*$ ($\because T = T^{-1}$)
$\Leftrightarrow$ T is unitary.

E16) a) Let $A = [a_{ij}]$, $B = [b_{ij}]$ and $AB = C = [c_{ij}]$. Then the (i, j)th element of

C^* = conjugate of the (j, i)th element of $C = \overline{\sum_{k=1}^{n} a_{jk} b_{ki}}$ ...... (1)

Now, if $B^* = [d_{ij}]$ and $A^* = [e_{ij}]$, then $d_{ij} = \overline{b_{ji}}$ and $\overline{e_{ij}} = a_{ji}$. Also,

$$\text{the (i, j)th element of } B^*A^* = \sum_{k=1}^{n} d_{ik} e_{kj} = \sum_{k=1}^{n} \overline{b_{ki}}\ \overline{a_{jk}}$$

$$= \sum_{k=1}^{n} \overline{a_{jk} b_{ki}} \qquad \text{...... (2)}$$

(1) and (2) $\Rightarrow C^* = B^*A^*$.

b) Let $B = A^{-1}$. Then $(AB)^* = I^* \Rightarrow B^*A^* = I$.
Similarly, $A^*B^* = I$.
$\therefore B^* = (A^*)^{-1}$, that is, $(A^{-1})^* = (A^*)^{-1}$.

E17) λ is a characteristic root of a Hermitian matrix A.
$\Leftrightarrow \lambda$ is an eigenvalue of A.
$\Leftrightarrow \lambda$ is an eigenvalue of A, treated as an operator.
$\Leftrightarrow \lambda$ is real, by Theorems 6 and 11.
$\therefore$, no characteristic root of A is purely imaginary.

E18) Let A be an upper triangular Hermitian matrix. Then $a_{ij} = 0$ for $i > j$. Also
$A = A^*$. $\therefore a_{ij} = \overline{a}_{ji}$.
$\therefore$, for $i < j$, $a_{ij} = \overline{a}_{ji} = \overline{0} = 0$, since $j > i$.
$\therefore \forall i \neq j$, $a_{ij} = 0$. $\therefore$ A is a diagonal matrix.
Similarly, if A is Hermitian and lower triangular, it must be a diagonal matrix.

E19) Let $B = \{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Let $[T]_B = A = [a_{ij}]$. Then
$[T^*]_B = A^* = [b_{ij}]$, $b_{ij} = \overline{a}_{ji}$. Now, $T = -T^* \Rightarrow [T]_B = -[T^*]_B$.
$\Rightarrow A = -A^*$.

E20) Since $\begin{bmatrix} 0 & 1 & i \\ -1 & i & 2 \end{bmatrix}$ is not a square matrix, it can't be unitary.

Now, if $A = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$, then $A^* = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}$.

$\therefore AA^* = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Similarly, $A^*A = I$.

$\therefore$ A is unitary.

If $A = \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix}$, then $A^* = \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix}$.

$\therefore AA^* = \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix} \begin{bmatrix} 1 & 1+i \\ 1-i & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1+i \\ 1-i & 2 \end{bmatrix} \neq I$.

$\therefore$ A is not unitary.

E21) Let $A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$. Then $A^* = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

$\therefore AA^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Also $A^*A = I$.

$\therefore$ A is orthogonal.

E22) The matrix is $A = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Then $A^* = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\therefore AA^* = A^*A = I.$

$\therefore$ A is unitary. $\therefore$, its columns form an orthonormal set of vectors.

E23) The eigenvectors corresponding to 6, −12 and 18 are $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$, respectively.

$\therefore$, the normalised eigenvectors are

$1/\sqrt{2}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, 1/\sqrt{6}\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, 1/\sqrt{3}\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$

$\therefore P = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 0 & 2/\sqrt{6} & -1/\sqrt{3} \end{bmatrix}$ is an orthogonal matrix such that

$P^t AP = \begin{bmatrix} 6 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & 18 \end{bmatrix}.$

UNIT 14 REAL QUADRATIC FORMS

Structure

14.1 INTRODUCTION

So far you have studied various kinds of matrices and inner products. In this unit we shall discuss a particular kind of inner product, which is closely connected to symmetric matrices. This is called a quadratic form. It can also be thought of as a particular kind of second degree polynomial, which is the way we shall first define it. We will discuss the geometric aspect of a particular case of quadratic forms in the next unit.

Quadratic forms are encountered in various mathematical and physical problems. For example, in physics, expressions for moment of inertia, energy, rate of generation of heat and stress ellipsoid in the theory of elasticity involve quadratic forms. Quadratic forms also appear while studying chemistry, the life sciences, and of course, many branches of mathematics.

In this unit we shall always assume that the underlying field is R.

Before going further make sure that you are familiar with Units 12 and 13.

Objectives

After reading this unit, you should be able to

- identify a real quadratic form;
- find the symmetric matrix associated to a quadratic form;
- calculate the rank of a quadratic form;
- obtain the orthogonal canonical reduction of a quadrstic form;
- find the normal canonical reduction of a quadratic form;
- calculate the signature of a quadratic form.

14.2 QUADRATIC FORMS

The word "quadratic" is not new to you. You have already encountered it when solving equations of the type

$ax^2 + bx + c = 0$, $a,b,c, \in \mathbf{R}$, $a \neq 0$, ... (1)

which are called quadratic equations. The left hand side of (1) is a quadratic function in one variable over **R**. We call the second degree term in (1), i.e., ax^2, **a quadratic form of order one**. It is called of order one, since it involves only one variable.

The most general quadratic equation over **R** involving two variables x and y is

$(ax^2 + 2hxy + by^2) + (2gx + 2fy) + c = 0$, $a,b,c,f,g,h \in \mathbf{R}$,

where at least one of a, h, b is non-zero. Its left hand side is a quadratic function, or quadratic polynomial, of order 2. The second degree terms occurring in this equation, i.e., the expression

$ax^2 + 2hxy + by^2$

is called a **quadratic form of order two**, since it involves two variables x and y.

The most general quadratic equation over **R** involving three variables is

$(ax^2 + by^2 + cz^2 + 2hxy + 2gxz + 2fyz) + 2ux + 2vy + 2wz + d = 0,$

a, b, c, d, f, g, h, u, v, w ∈ **R**, where at least one of a, b, c, f, g, h is non-zero. Its left hand side is a quadratic function, or quadratic polynomial, in three variables. The bracketed part of this equation, containing only second degree terms, is called a **quadratic form of order three.**

By now you can see how we can generalise this concept. We call the non-zero form

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j$$

A polynomial is called homogeneous if each of its terms has the same degree.

a quadratic form over **R** of order n, where the a_{ij}'s are real constants and $x_1, x_2, \ldots, x_n$ are real variables.

Note: These expressions are called quadratic, since they are of second degree. They are called forms, since every term in them has the same degree.

We are now ready to make a formal definition.

Definition: A homogeneous polynomial of degree two is called a **quadratic form**. Its **order** is the number of variables that occur in it.

For example, $x^2 - 3y^2 + 4xz$ is a quadratic form of order 3.

A quadratic form is real if its variables can only take real values and the coefficients are real numbers. We have already stated, in the unit introduction, that all spaces considered in this unit shall be over **R**. **Therefore, by a quadratic form we shall always mean a real quadratic form.**

From the definition of a quadratic form it is clear that a real valued function will be a quadratic form if and only if it satisfies each of the following conditions:

a) it is a polynomial,

b) it is homogeneous, and

c) it is of degree two.

Let us look at some examples now.

Example 1: Which of the following are quadratic forms? In the case of quadratic forms, find the order.

a) $x^2 + x + 1$

b) $2x^2 + y^2 + z^2$

c) $x^2 - \sqrt{2}y^2 = 0$

d) $3x_1^2 + x_1 x_2 - \sqrt{3}x_2^2$

e) $x_1^3 - x_2^3 + x_2^2 x_1$

f) $x^3 + x^2y - y^3$

g) $x^2 + \log x$.

Solutions: (c) is an equation, and not a polynomial. (a) and (e) are polynomials, but they are not homogeneous. (f) is a polynomial which is homogeneous, but its degree is three and not two. (g) is not a polynomial. Only (b) and (d) represent quadratic forms. (b) involves three variables, and hence, its order is three. (d) involves two variables, and thus, has order two.

Try the following exercises now.

E E 1) Give an example of a function that is

a) a non-homogeneous polynomial of degree 2.

b) a homogeneous polynomial, but not of degree 2.

E E 2) Which of the following represent quadratic forms?

a) $x^2 - xy$

b) $x_1 + x_2$

c) x_1^3

d) $x^3 - xy^2$

e) $\sin(x^2 + 2y^2)$

f) $x_1^2 - \sqrt{2}\, x_2^2 = 0$

E 3) Find the values of the integer k for which the following will represent quadratic forms.

a) $x^2 - 2y^2 - kxy^2$

b) $x^k + 2y^2$

c) $x_1^4 + 2x_1x_2 - x_1^k$

E 4) Let Q_1 and Q_2 be two quadratic forms, both of order n, in the n variables $x_1, x_2, \ldots, x_n$. Which of the following will be a quadratic form?

$Q_1 + Q_2, aQ_1 + bQ_2, Q_1 - Q_2, Q_1Q_2, Q_1/Q_2$.

Let us now see how to represent a quadratic form as a product of matrices. In fact, you will see how a quadratic form can be written as an inner product.

14.3 QUADRATIC FORM AS MATRIX PRODUCT

Consider the quadratic form of order two,

$Q = 2x^2 + 2xy + 3y^2$.

Putting $X = \begin{bmatrix} x \\ y \end{bmatrix}$ and $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$, we find that

$$Q = X^tAX = [x\ y] \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \qquad \ldots\ldots (1)$$

The question now is whether we can replace the matrix A by another matrix without changing the quadratic form Q. In fact, you can check that

$Q = X^tBX$, where $B = \begin{bmatrix} 2 & 2 \\ 0 & 3 \end{bmatrix}$, and

$Q = X^tCX$, where $C = \begin{bmatrix} 2 & -1 \\ 3 & 3 \end{bmatrix}$.

Thus, we see that if we replace A by B or C in (1), the quadratic form is not changed. This shows us that the choice of the matrix A in (1) is not unique. In this section we shall find the reason for this, and also investigate the general matrix which can replace A in (1).

Note that we can also write $Q = \langle AX, X \rangle$, where $\langle Y, Z \rangle = Z^tY$ for any $Y, Z \in V_2(\mathbf{R})$. So, as you go along, remember that we are simultaneously discussing the representation of Q as a matrix product, as well as an inner product.

Look carefully at the matrices A, B and C, given above. Do they have a common feature? You must have noticed that the diagonal elements of all these matrices are the same, i.e., A, B and C have the same diagonal. Now, what about the off-diagonal (i.e., non-diagonal) entries? Have you noticed that the sum of the off-diagonal entries in all these matrices is 2? Note that the coefficient of the term xy, of the given quadratic form, is also 2.

E 5) Change one of the diagonal entries of A and verify that this will change the quadratic form.

In fact, any matrix $P = \begin{bmatrix} 2 & a \\ b & 3 \end{bmatrix}$, with a + b = 2, can replace A without changing the quadratic form Q. This is because the coefficient of xy in the quadratic form X^tPX is (a + b). However, if we insist that the matrix P should be symmetric, then we must have a = b; and hence, the choice is unique, namely, $\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$.

We, therefore, conclude that A is the only symmetric matrix for which $Q = X^tAX$.

This symmetric matrix A is called the **matrix of the quadratic form Q, or the matrix associated to the quadratic form Q.** Observe that

$$A = \begin{bmatrix} \text{coef. of } x^2 & (1/2)\text{ coef. of } xy \\ (1/2)\text{ coef. of } xy & \text{coef. of } y^2 \end{bmatrix},$$

where coef. is short for coefficient.

We can sum up the above discussion as follows:

Given a quadratic form Q of order 2, there are infinitely many square matrices B for which $Q = X^tBX$. However, there will be a **unique symmetric matrix** A for which $Q = X^tAX$. This matrix A, which is called the matrix of the quadratic form Q, is given by the rule

$$A = \begin{bmatrix} \text{coef. of } x^2 & (1/2)\text{ coef. of } xy \\ (1/2)\text{ coef. of } xy & \text{coef. of } y^2 \end{bmatrix} \quad \text{........(2)}$$

Actually, there is a one-to-one correspondence between the set of all symmetric square matrices of order 2 and the set of all quadratic forms of order 2. This is because, given any 2 × 2 symmetric matrix $B = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, we can obtain a unique quadratic form of order 2 corresponding to it, namely, $X^tBX = ax^2 + 2bxy + dy^2$. Conversely, given any quadratic form of order 2 we can obtain a unique 2 × 2 symmetric matrix by the rule (2). The following examples will illustrate this correspondence.

Example 2: What is the quadratic form generated by

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}?$$

Solution: The quadratic form generated by A is

$$[x\ y]\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

On expanding this we get

$x^2 - 2xy + y^2$

Observe that you could have obtained the quadratic form simply by applying the rule (2) as follows:

Comparing the given matrix A with the matrix in (2) gives

coef. of $x^2 = 1$, coef. of $y^2 = 1$, (1/2) coef. of xy = −1.

Therefore, the required quadratic form is $x^2 - 2xy + y^2$.

Example 3: A general diagonal matrix of order 2 is $A = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$.

What is the corresponding quadratic form?

Solution: Once again you can either compute

$$X^tAX = [x\ y]\begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \alpha_1 x^2 + \alpha_2 y^2.$$

The matrix of a diagonal form is a diagonal matrix.

or use rule (2) to get

coef. of $x^2 = \alpha_1$, coef. of $y^2 = \alpha_2$, coef. of xy = 0.

∴ the required form is $\alpha_1 x^2 + \alpha_2 y^2$.

Such a quadratic form is called a diagonal form.

Example 4: Find the matrices associated to the following quadratic forms.

a) x^2

b) $-y^2-4xy$

Solution: Rule (2) is very handy for writing the symmetric matrix of a given quadratic form. It is easy to see that the corresponding matrices will be

a) $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, b) $\begin{bmatrix} 0 & -2 \\ -2 & -1 \end{bmatrix}$.

Now for an exercise!

E 6). Find the 2×2 matrices associated to

a) $-y^2$, b) $2x^2+y^2$, c) $2xy$, d) $px^2+qxy+ry^2$.

The above discussion involved matrices and quadratic forms of order two. It can be extended to matrices and quadratic forms of higher orders. Let us look at the case of quadratic forms of order 3.

Let us consider a general 3×3 matrix

$$A=[a_{ij}]=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The quadratic form determined by A will be

$$Q=X^tAX, \qquad \ldots\ldots (3)$$

where $X^t=[x_1 \ x_2 \ x_3]$.

Expand the matrix product in (3) and verify that

$$Q=a_{11}x_1^2+a_{22}x_2^2+a_{33}x_3^2+(a_{12}+a_{21})x_1x_2+(a_{23}+a_{32})x_2x_3+(a_{13}+a_{31})x_1x_3. \quad \ldots\ldots (4)$$

Observe that the diagonal elements of A, i.e., a_{11}, a_{22} and a_{33}, are the coefficients of x_1^2, x_2^2, and x_3^2, respectively, in Q given by (4).

Also note that the sum of the two entries a_{12} and a_{21} determines the coefficient of x_1x_2, while these two entries do not occur elsewhere in (4). So, if we replace a_{12} and a_{21} by two different numbers a'_{12} and a'_{21} such that $a'_{12}+a'_{21}=a_{12}+a_{21}$, while keeping other entries of A unchanged, the new matrix A′, thus obtained, will not be equal to A. But the quadratic forms generated by A and A′ will be the same, i.e.,

$$Q=X^tAX=X^tA'X.$$

Similar changes can be made for the entries contributing to the coefficients of x_1x_3, to obtain matrices different from A which can replace A without changing the quadratic form.

However, if the matrix A′ is restricted to being symmetric then the choice is unique, i.e.,

$$a'_{12}=a'_{21}=\frac{1}{2}(a_{12}+a_{21})=\frac{1}{2}\text{ (coef. of } x_1x_2),$$

$$a'_{13}=a'_{31}=\frac{1}{2}(a_{13}+a_{31})=\frac{1}{2}\text{ (coef. of } x_1x_3),$$

and $a'_{23}=a'_{32}=\frac{1}{2}(a_{23}+a_{32})=\frac{1}{2}\text{ (coef. of } x_2x_3).$

Therefore, the unique symmetric matrix corresponding to the quadratic form (4) will be

$$A' = \begin{bmatrix} \text{coef. of } x_1^2 & \frac{1}{2}\text{ coef. of } x_1x_2 & \frac{1}{2}\text{ coef. of } x_1x_3 \\ \frac{1}{2}\text{ coef. of } x_1x_2 & \text{coef. of } x_2^2 & \frac{1}{2}\text{ coef. of } x_2x_3 \\ \frac{1}{2}\text{ coef. of } x_1x_3 & \frac{1}{2}\text{ coef. of } x_2x_3 & \text{coef. of } x_3^2 \end{bmatrix} \quad \ldots\ldots (5)$$

We sum up the above discussion as follows:

Given a quadratic form of order 3, there are infinitely many matrices of order 3 which will generate it. However, a symmetric matrix that will generate a quadratic form of order three is unique. This symmetric matrix is called the matrix associated to the quadratic form, or simply, the matrix of the quadratic form.

Just as in the case of order 2 forms, there is a one-to-one correspondence between the set of all symmetric matrices of order three and the set of all quadratic forms of order three. The next few examples will illustrate the above discussion.

Example 5: Find the quadratic form Q corresponding to the symmetric matrix

$$A = \begin{bmatrix} 1 & -2 & 3 \\ -2 & 4 & 1 \\ 3 & 1 & 2 \end{bmatrix} = [a_{ij}], \text{ say.}$$

Solution: A straight-forward way will be to expand X^tAX where $X^t = [x_1, x_2, x_3]$. Then we would get

$$Q = x_1^2 + 4x_2^2 + 2x_3^2 - 4x_1x_2 + 6x_1x_3 + 2x_2x_3.$$

But, a quicker way is to use the rule (5). Comparing the entries of A′ in (5) with those of A above we can obtain all the coefficients of the quadratic form as follows:

Coefficients of x_1^2, x_2^2, x_3^2 will be the elements of the diagonal in A, i.e., 1, 4 and 2, respectively.

coef. of $x_1x_2 = a_{12} + a_{21} = -4$

coef. of $x_1x_3 = a_{13} + a_{31} = 6$

coef. of $x_2x_3 = a_{23} + a_{32} = 2$

Then the required quadratic form is Q, as obtained above.

Example 6: Find the symmetric matrix associated with the form

$2x_1^2 - x_2^2 + x_3^2 + 2x_1x_2 - 6x_1x_3$.

Solution: Using the rule (5), we can write the matrix as

$$\begin{bmatrix} 2 & 1 & -3 \\ 1 & -1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$$

Example 7: Find the quadratic form associated with the zero matrix of order three.

Solution: All the entries of a zero matrix are zero. Therefore, using (5), we get all the coefficients to be zero. The associated quadratic form is, then,

$0x_1^2 + 0x_2^2 + 0x_3^2 + 0x_1x_2 + 0x_1x_3 + 0x_2x_3$,

which is the zero quadratic form of order three.

Example 8: Consider the general diagonal matrix of order three,

$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$ What is the associated quadratic form?

Solution: The associated quadratic form is the diagonal form

$\lambda_1x_1^2 + \lambda_2x_2^2 + \lambda_3x_3^2$.

The following exercises deal with quadratic forms of orders 2 and 3.

E E7) Write the following quadratic forms as X^tAX, where A is a symmetric matrix.

a) $7x^2 + 7y^2 - 2z^2 + 20yz - 20zx - 2xy$ (in $\mathbf{R}^3$)

b) $x_1^2 + x_2^2 - x_1x_2$ (in $\mathbf{R}^2$)

c) $x_1^2 - 2x_1x_2$ (in $\mathbf{R}^3$)

d) $2yz + 2zx$ (in $\mathbf{R}^3$)

E8) Expand X'AX as a polynomial, where $X^t = [x, y, z]$, and A is

a) $\begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix}$, b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, c) $\begin{bmatrix} 4 & 0 & 0 \\ 0 & -\sqrt{2} & 0 \\ 0 & 0 & -1 \end{bmatrix}$.

Can we extend the comments about quadratic forms of order two and three to a quadratic form of any finite order n? Yes. You know that a general quadratic form of order n is given by

$$Q = \sum_{i,j=1}^{n} a_{ij}x_ix_j, \text{ where } a_{ij} = a_{ji} \forall i, j = 1, \ldots, n.$$

The associated symmetric matrix A of order n will be

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}, \text{ where } a_{ij} = a_{ji} \forall i, j = 1, \ldots, n.$$

Thus, Q can be written as

$A = X^tAX$, where $X^t = [x_1\ x_2 \ldots x_n]$.

So, there is a one-to-one correspondence between the set of all symmetric matrices of order n and the set of quadratic forms of order n. Under this correspondence the matrix A corresponds to the quadratic form $X^t AX$. The following exercise illustrates this for order 4.

E9) Expand X^tAX as a polynomial, where $X^t = [x_1, x_2, x_3, x_4]$ and

$$A = \begin{bmatrix} 4 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 \\ 3 & 6 & 0 & 1 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

Find the symmetric matrix A' such that $X^tAX = X^tA'X$.

Before going further, we would like to remind you that the quadratic form of order n, X^tAX, is simply the inner product $\langle AX, X\rangle$ in $V_n(R)$.

Let us now see what happens to the matrix of a quadratic form if we change the basis of the underlying vector space.

14.4 TRANSFORMATION OF A QUADRATIC FORM UNDER A CHANGE OF BASIS

In the previous section you have seen that a quadratic form Q of order n can be expressed as X^tAX, where $X^t = [x_1, x_2, \ldots, x_n]$ and A is a real symmetric matrix of order n. Now, $x_1, x_2, \ldots, x_n$ are the components (or the coordinates) of the vector X with respect to a pre-assigned basis $\{e_1, e_2, \ldots, e_n\}$ of R^n. If we change the basis of R^n from $B = \{e_1, e_2, \ldots, e_n\}$ to another basis $B' = \{e_1', \ldots, e_n'\}$, the components of X will also change. Therefore, the quadratic form Q will also change. We will show that, under a change of basis, the quadratic form changes according to a certain transformation law.

Let P be the matrix of the change of basis from B to B' (see Sec. 7.6). Then $P = [a_{ij}]$,

where $e_j' = \sum_{i=1}^{n} a_{ij} e_i$.

You have seen, in Unit 7, that P is invertible. Note that the columns of P are the components of the vectors of the new basis B', expressed in terms of the original basis B.

Now, if $X^t = [x_1, \ldots, x_n]$ and $Y^t = [y_1, \ldots, y_n]$ denote the coordinates of a vector in R^n with respect to B and B', respectively, then

$$\sum_{i=1}^{n} x_i e_i = \sum_{j=1}^{n} y_j e_j' = \sum_{i}\sum_{j=1}^{n} y_j a_{ij} e_i.$$

Since $\{e_1, \ldots, e_n\}$ is a basis, we get

$$x_i = \sum_{j=1}^{n} a_{ij} y_j \ \forall\ i = 1, 2, \ldots, n.$$

This is equivalent to the matrix equation

$$\begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11} & \ldots\ldots & a_{1n} \\ \cdot & \ldots\ldots & \cdot \\ \cdot & \ldots\ldots & \cdot \\ a_{n1} & \ldots\ldots & a_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

i.e., $X = PY$.

This equation is the coordinate transformation corresponding to the change of basis from B to B′. The change of basis will convert the quadratic form X^tAX into

$$(PY)^t A(PY) = Y^t(P^tAP)Y$$
$$= Y^tCY, \text{ where } C = P^tAP.$$

But, is C symmetric? Well, $C^t = (P^tAP)^t = P^tAP = C$.

∴ C is symmetric.

The above discussion shows that, under a change of basis given by the invertible matrix P, the coordinate transformation is given by $X = PY$, and the quadratic form X^tAX gets transformed into another quadratic form Y^tCY, where $C = P^tAP$. This leads us to the following definitions.

Definitions: Two real symmetric matrices A and B are called **congruent** if there exists an invertible real matrix P such that $B = P^tAP$.

Two quadratic forms X^tAX and Y^tBY are called **equivalent** if their matrices, A and B, are congruent.

In particular, if the matrices A and B are orthogonally similar (see Unit 13) then the corresponding quadratic forms, X^tAX and Y^tBY are called **orthogonally equivalent.**

So, under a change of basis, a quadratic form gets transformed to an equivalent quadratic form. They may or may not be orthogonally equivalent. Let us look at an example.

Example 9: Consider the change of basis of $\mathbf{R}^2$ from the standard basis $B_1 = \{(1, 0), (0, 1)\}$ to $B_2 = \{(1, 0), (1, 2)\}$. Let (x_1, x_2) and (y_1, y_2) represent coordinates with respect to B_1 and B_2, respectively.

a) Find the coordinate transformation that expresses x_1, x_2 in terms of y_1, y_2.

b) Let $Q(X) = x_1^2 - 2x_1x_2 + 4x_2^2$. Find the expression of Q in terms of y_1 and y_2.

Solution: a) The change of basis from B_1 to B_2 is given by the coordinate transformation.

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \text{ or } X = PY, \text{ say.} \quad \ldots\ldots (1)$$

(Remember that the columns of P will be the components of the new basis vectors expressed in terms of the old basis.) From (1)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} y_1 + y_2 \\ 2y_2 \end{bmatrix}$$

i.e., $x_1 = y_1 + y_2$

$x_2 = 2y_2$,

which is the required coordinate transformation.

b) Now $Q(X) = [x_1, \; x_2]\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$= X^tAX, \text{ say} \quad \ldots\ldots (2)$$

Using (1), $Q(Y) = Y^t(P^tAP)\,Y$. ...... (3)

where

$$P^tAP = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & -1 \\ -1 & 13 \end{bmatrix}$$

Using this in (3), we get

$$Q(Y) = y_1^2 - 2y_1y_2 + 13y_2^2 \quad \ldots\ldots (4)$$

Thus, under the change of basis given by $X = PY$, the given quadratic form transforms into (4).

The following exercises will give you some more practice in dealing with quadratic forms under a change of basis.

$(AB)^t = B^tA^t$
$(A^t)^t = A$

E10) Verify that the matrix P in Example 9 is not orthogonal. (Therefore, (1) is not an orthogonal transformation. Therefore, (2) and (4) are equivalent, but not orthogonally equivalent.)

E11) Consider the quadratic form given in Example 9. Replace B_2 by {(1, 0), (1, 1)}. Is this change of basis orthogonal? Find the quadratic form with respect to the new basis B_2.

E12) Let a quadratic form have expression

$7x^2 + 52xy - 32y^2$,

with respect to the standard basis B_1 = {(1, 0), (0, 1)} of R^2. Find its expression with respect to the basis B_2 = {(2, 1), (1, −2)}.

E13) Consider the change of basis of R^3 from the standard basis B = {(1, 0, 0), (0, 1, 0), (0, 0, 1)} to the basis B′ = {(−2, 6, 3), (3, −2, 6), (6, 3, −2)}. Find the coordinate transformation corresponding to this change of basis.

Now let us see what we mean by the rank of a quadratic form.

14.5 RANK OF A QUADRATIC FORM

In Unit 8 you have studied about the rank of a matrix. Here we will discuss the rank of a quadratic form. Since quadratic forms are closely associated with matrices, the concept of the rank of a matrix can be used to define the rank of a quadratic form. But first we shall prove the following result.

Theorem 1: Congruent matrices have the same rank.

Proof: Let A and B be congruent matrices. Then there is a non-singular matrix P such that $B = P^t AP$.

Recall, from Unit 8, that multiplication by a non-singular matrix does not change the rank of a matrix. Therefore,

rank (B) = rank (P^tAP) = rank (A),

which proves the theorem.

We are now all set to define the rank of a quadratic form.

Definition: The **rank of a quadratic form** is the rank of its associated matrix.

You may think that this definition is not meaningful, because the associated matrix depends on the basis of the vector space. But Theorem 2 assures us that the definition is meaningful.

Theorem 2: The rank of a quadratic form does not change under a change of basis.

Proof: Let $Q(X) = X^tAX$ be a quadratic form of rank r. Under a change of basis let X = PY. Then $Q(Y) = Y^t(P^tAP)Y$.

And then, rank Q(X) = rank A = rank (P^tAP) (by Theorem 1)

= rank Q(Y)

Thus, we have proved the theorem.

Try the following simple exercise.

E E14) Verify that the rank of a diagonal form is the number of non-zero terms in its expression.

Now let us obtain the ranks of some more quadratic forms.

Example 10: Consider the quadratic form

$$Q(X) = 2x_1^2 + 2x_1x_2 + 2x_2^2 - 6x_1x_3 - 6x_2x_3 + 6x_3^2,$$

where $[x_1, x_2, x_3]$ are the coordinates of X with respect to the standard basis of $\mathbf{R}^3$.

a) Find the expression of Q with respect to the basis

$$B = \left\{\left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}\right), \left(\frac{1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)\right\}$$

b) What is the rank of Q?

Solutions: a) Let $Y^t = [y_1, y_2, y_3]$ denote the coordinates with respect to the new basis B. Then, the change of coordinates is given by

$$X = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} Y = PY \ \text{(say)}$$

The given quadratic form can be written as X^tAX, where

$$A = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 2 & -3 \\ -3 & -3 & 6 \end{bmatrix}$$

The change of coordinates given by $X = PY$ will convert X^tAX into $Y^t(P^tAP)Y$, where

$$P^tAP = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} 2 & 1 & -3 \\ 1 & 2 & -3 \\ -3 & -3 & 6 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{-2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} = \begin{bmatrix} 9 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using this, we get

$$Q(Y) = 9y_1^2 + y_2^2,$$

which is the required quadratic form.

Note that P is an orthogonal matrix. $\therefore$ **Q(X) and Q(Y) are orthogonally equivalent.**

b) Now, let us obtain rank(Q) directly. We know that rank (A) = 2.

$\therefore$ rank $(X^tAX) = 2$, i.e., the rank of Q is 2.

Another way of showing that rank Q(X) = 2 is as follows: Q(X) and Q(Y) are equivalent, and the rank of the diagonal quadratic form Q(Y) is two. $\therefore$, rank of Q(X) is also two.

The following exercise will give you some practice in obtaining the rank of a quadratic form.

E15) Find the rank of the following quadratic forms in $\mathbf{R}^3$.

a) $5x^2 + 6y^2 + 7z^2 - 4xy - 4yz$

b) $x^2 + y^2 + z^2 + 2xy + 2yz + 2xz$

c) $2x^2 + 2y^2 + 2z^2 + 2xy + 2yz + 2xz$

d) $x^2 - y^2$

Now, as we have seen, under a change of basis a quadratic form gets transformed to an equivalent quadratic form. We will show that all quadratic forms can be divided into equivalence classes based on the relationship between their matrices. Recall from Unit 1 that a relation is an equivalence relation if and only if it is reflexive, symmetric and transitive.

E E16) Recall the definition of congruent and orthogonally similar matrices. Show that the relations of congruence and orthogonal similarity between matrices are equivalence relations.

Once you have proved E16 the following theorem follows immediately.

Theorem 3: The relation of equivalence, as well as orthogonal equivalence, of quadratic forms is an equivalence relation.

Proof: We will prove the theorem for equivalence. You can prove the result for orthogonal equivalence similarly.

Now two quadratic forms X^tAX and Y^tBY are equivalent if and only if A and B are congruent. You have just proved (in E16) that the congruence of matrices is an equivalence relation. ..., the equivalence of quadratic forms is also an equivalence relation.

In view of Theorem 3, the relation of equivalence (respectively, orthogonal equivalence) divides the set of all quadratic forms of order n into disjoint equivalence classes. Each equivalence class contains all quadratic forms which are equivalent (respectively, orthogonally equivalent) to each other. In other words, any two quadratic forms in an equivalence class can be obtained from each other by a suitable change of basis. This division into classes will be very useful in the next unit.

We shall now use results of Units 12 and 13 to establish a method to reduce a quadratic form into a diagonal form, by using a suitable orthogonal change of basis.

14.6 ORTHOGONAL CANONICAL REDUCTION

Recall from Unit 13 that for any real symmetric matrix A, we can always construct an orthogonal matrix R whose columns are a set of orthonormal eigenvectors (say, $U_1, U_2, \ldots, U_n$) of A such that

$R^tAR = \text{diag}(\lambda_1, \ldots, \lambda_n)$, ... (1)

$\lambda_1, \ldots, \lambda_n$ being the eigenvalues of A corresponding to the eigenvectors $U_1, \ldots, U_n$, respectively.

Remember, R may not be unique. This could be due to two factors:

i) Changing the order in which eigenvectors are taken will change R.

ii) An orthonormal eigenvector corresponding to an eigenvalue need not be unique.

We shall now use the relation (1) to transform any quadratic form to a diagonal form.

Let A be the matrix of a quadratic form with respect to a pre-assigned basis. Let R be an orthogonal matrix obtained from A as indicated above. Now consider the change of basis from the pre-assigned basis to the basis $\{U_1, U_2, \ldots, U_n\}$. The coordinate transformation will be given by

$X = RY$, ... (2)

$Y^t = [y_1, y_2, \ldots, y_n]$ being the coordinates with respect to the new basis. R being orthogonal, (2) is an orthogonal transformation which will convert X^tAX into

$Y^t(R^tAR)Y = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2$, ... (3)

because of (1).

Thus X^tAX is orthogonally equivalent to the diagonal form in (3) whose coefficients are the eigenvalues of A. The form in (3) is called an **orthogonal canonical reduction of X^tAX.**

We say that the orthogonal transformation (2) has **reduced the quadratic form X^tAX into its orthogonal canonical form,** given by (3). The form in (3) is orthogonal since the transformation used to convert X^tAX into it is orthogonal. It is called canonical as the reduced form is the simplest orthogonal reduction of X^tAX. The elements of the basis which diagonalise the quadratic form (in this case they are $U_1, \ldots, U_n$) are called the **principal axes** of the quadratic form. In Unit 15 you will realise why they are called axes.

We can summarise the above discussion in the form of a theorem.

Theorem 4: A real quadratic form X^tAX can always be reduced to the diagonal form

$$\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2$$

by an orthogonal change of basis, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. The new ordered basis is an orthonormal set of eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$.

Now, if the matrix of a quadratic form is orthogonally similar to diag $(\lambda_1, \ldots, \lambda_n)$, it is also orthogonally similar to diag $(\lambda_2, \lambda_1, \ldots, \lambda_n)$. Thus, the orthogonal canonical form to which a quadratic form is orthogonally equivalent is unique except for the order of the coefficients. If we insist that the non-zero eigenvalues be written in decreasing order followed by the zero eigenvalues, if any, then we can obtain a unique orthogonal canonical form.

So, we can state the following result.

Theorem 5: A quadratic form of rank r is orthogonally equivalent to a unique orthogonal canonical form $\lambda_1 y_1^2 + ... + \lambda_r y_r^2$, where $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of the matrix of the quadratic form, such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$.

Proof: Let X^tAX be a quadratic form of rank r. Then rank (A) = r. Therefore, A has r non-zero eigenvalues. We write them as $\lambda_1, \ldots, \lambda_r$, in decreasing order. Now, by Theorem 4 we get the required result.

So far we have spoken about the orthogonal canonical form in an abstract way. Let us now look at a practical method of reducing a quadratic form to its orthogonal canonical form.

Step by step procedure for orthogonal canonical reduction: We will now give the sequence of operations which are needed to reduce a given quadratic form to its orthogonal canonical form, and to obtain the required coordinate transformations or the new basis.

1) Construct the symmetric matrix A associated to the given quadratic form $\sum_{i,j=1}^{n} a_{ij}x_i x_j$.
2) Form the characteristic equation

   $\det(A - \lambda I) = 0$

   and find the eigenvalues of A. Let $\lambda_1, \ldots, \lambda_r$ be the non-zero eigenvalues arranged in decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r$.
3) An orthogonal canonical reduction of the given quadratic form is

   $\lambda_1 y_1^2 + \ldots + \lambda_r y_r^2$.
4) Obtain an ordered system of n orthonormal vectors $U_1, \ldots, U_n$ consisting of eigenvectors corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_n$ (here $\lambda_{r+1} = 0 = \ldots = \lambda_n$). Note that for **repeated eigenvalues also we must obtain linearly independent orthonormal eigenvectors.**
5) Construct the orthogonal matrix P whose columns are the eigenvectors $U_1, \ldots, U_n$.
6) The required change of basis is given by X = PY.
7) The new basis $\{U_1, U_2, \ldots, U_n\}$ is called the canonical basis and its elements are the principal axes of the given quadratic form.

In Step 2 you are required to find the eigenvalues, i.e., the roots of the characteristic equation. In a realistic situation the roots can be irrational numbers and we may have to use numerical methods to determine such roots. We have avoided irrational numbers by carefully selecting the quadratic forms in our examples and exercises so that the roots of characteristic equations are rational numbers.

To clarify the procedure given above we present some examples and exercises.

Example 11: Obtain the unique orthogonal canonical form of the quadratic form

$$5x_1^2 - 6x_1x_2 + 5x_2^2.$$

Also give the associated coordinate transformation, canonical basis and principal axes of the given form.

Solution: The matrix of this quadratic form is

$$A = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$$

The eigenvalues of A are given by

$$\begin{vmatrix} \lambda - 5 & 3 \\ 3 & \lambda - 5 \end{vmatrix} = 0,$$

i.e., $\lambda^2 - 10\lambda + 16 = 0 \Rightarrow \lambda = 8, 2$.

Thus, the required orthogonal canonical reduction will be

$$8y_1^2 + 2y_2^2.$$

The normalised eigenvectors corresponding to the eigenvalues 8 and 2 are U_1 and U_2, where

$$U_1 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \text{ and } U_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

Thus, the new orthonormal basis is $\{U_1, U_2\}$, which is the canonical basis. U_1 and U_2 are the principal axes of the given form.

The associated coordinate transformation will be

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

i.e., $x_1 = 1/\sqrt{2}(-y_1 + y_2)$

$x_2 = 1/\sqrt{2}(y_1 + y_2)$

Note: Remember that the choice of normalised eigenvectors is not unique. You could have as well taken $-U_1$ or $-U_2$, instead of U_1 and U_2, respectively.

E17) In Example 11 take the normalised eigenvectors corresponding to 8 and 2 to be $-U_1$ and $-U_2$, respectively. Find the coordinate transformation needed for the orthogonal canonical reduction.

Now we look at an example in which the associated matrix has repeated eigenvalues.

Example 12: Consider the quadratic form

$x^2 + y^2 + z^2 + 2xy + 2xz + 2yz$ ...... (1)

Find its orthogonal canonical reduction and the corresponding new basis.

Solution: The matrix of (1) is

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

The eigenvalues of A are 3, 0, 0. Thus, the orthogonal canonical reduction of (1) is

$3x_1^2$, ... (2)

where x_1, y_1, z_1 are the new coordinates.

A normalised eigenvector corresponding to the eigenvalue 3 is

$$\begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}.$$

Eigenvectors corresponding to the eigenvalue 0 are given by

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

i.e., $x + y + z = 0$ ...... (3)

Here we can choose any two mutually orthogonal normalised vectors satisfying (3). Let us choose

$$\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ -2/\sqrt{6} \end{bmatrix}.$$

The new basis, in this case, is

$$\left\{\begin{bmatrix}1/\sqrt{3}\\1/\sqrt{3}\\1/\sqrt{3}\end{bmatrix}, \begin{bmatrix}1/\sqrt{2}\\-1/\sqrt{2}\\0\end{bmatrix}, \begin{bmatrix}1/\sqrt{6}\\1/\sqrt{6}\\-2/\sqrt{6}\end{bmatrix}\right\},$$

which is the canonical basis. Its elements are the principal axes of (1). The change of basis needed to convert (1) into (2) is given by

$$\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6}\\1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6}\\1/\sqrt{3} & 0 & -2/\sqrt{6}\end{bmatrix}\begin{bmatrix}x_1\\y_1\\z_1\end{bmatrix}.$$

We again observe that the canonical basis, principal axes and the coordinate transformation needed for reduction are not uniquely determined. We could have chosen any two mutually orthogonal orthonormal eigenvectors of 0.

The next few exercises will give you some practice in applying the procedure of reduction.

E E18) Find the orthogonal canonical forms to which the following quadratic forms can be reduced by means of an orthogonal change of basis. Also obtain a set of principal axes for them.

a) $x^2 + 4xy + y^2$

b) $8x^2 - 4xy + 5y^2$

c) $3x_2^2 + 3x_3^2 + 4x_1x_2 + 4x_1x_3 - 2x_2x_3$

E19) Which of the following quadratic forms are orthogonally equivalent?

a) $9x_2^2 + 9x_3^2 + 12x_1x_2 + 12x_1x_3 - 6x_2x_3$

b) $-3y_1^2 + 6y_2^2 + 6y_3^2 - 12y_1y_2 + 12y_1y_3 + 6y_2y_3$

c) $11z_1^2 - 4z_2^2 + 11z_3^2 + 8z_1z_2 - 2z_1z_3 + 8z_2z_3$

E20) Show that the quadratic forms

$x^2 - 2y^2 + z^2$ and $z_1^2 - 2x_1^2 + y_1^2$

are orthogonally equivalent. Find the orthogonal transformation which will transform the first of these into the second.

We will now try to reduce the matrix of a quadratic form to a diagonal form whose diagonal elements are only 1, −1 or 0.

14.7 NORMAL CANONICAL FORM

If we do not restrict ourselves to an orthogonal change of basis, then we can reduce a quadratic form to a simpler form than the one we considered in the previous section. In this simpler version the coefficients of the reduced form are ± 1 or zero.

Let $X^t AX = \sum_{i,j=1}^{n} a_{ij} x_i x_j$ ...... (1)

be a quadratic form of order n. From Theorem 5 we know that $X^t AX$ can be reduced to its unique orthogonal canonical form

$\lambda_1 y_1^2 + \ldots\ldots + \lambda_r y_r^2$, ...(2)

where $\lambda_1, \ldots\ldots, \lambda_r$ are the non-zero eigenvalues of A such that $\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_r$. Thus, rank (A) = r or, equivalently, the rank of (1) is r.

Now consider the coordinate transformation

$$\left.\begin{array}{l} z_i = \sqrt{|\lambda_i|}\, y_i,\ i = 1, 2, \ldots\ldots, r \\ z_i = y_i,\ i = r+1, \ldots\ldots, n \end{array}\right] \quad \ldots\ldots (3)$$

for $a \in R$

$$\text{sign}(a) = \begin{cases} 1, \text{ if } a > 0 \\ -1, \text{ if } a < 0 \\ 0, \text{ if } a = 0 \end{cases}$$

This is a non-singular transformation which will convert (2) into

$\text{sign}(\lambda_1) z_1^2 + \ldots\ldots + \text{sign}\ (\lambda_n) z_n^2$ ...... (4)

i.e., $\sum_{k=1}^{n} \text{sign}\ (\lambda_i) z_i^2$

Remember, sign $(\lambda_{r+1}) = 0 = \ldots = \text{sign}\ (\lambda_n)$.

Thus, by two successive transformations, one orthogonal and the other non-singular, we have reduced the given quadratic form to a diagonal form (4) of order n whose coefficients are $\pm$ 1 or 0. We call the form (4) the normal canonical form of the quadratic form (1). We give the following definition.

Definition: A diagonal quadratic form, whose coefficients are $\pm$ 1 or 0, is called a **normal canonical form.**

For example, $x^2 - y^2$ is a normal canonical form, but $2x^2 + y^2$ is not.

The procedure involved in transforming (1) to (4) is described as **reducing a quadratic form to its normal canonical form.**

E21) The transformation (3) is not, in general, an orthogonal transformation. Under what conditions will it become orthogonal?

We can sum up the above discussion in the following theorem.

Theorem 6: A real quadratic form can always be reduced to a normal canonical form by a suitable non-singular transformation.

Let us now look at some examples that will help you in understanding the procedure.

Example 13: Reduce the quadratic form

$5x_1^2 - 6x_1x_2 + 5x_2^2$ ...... (1)

to a normal canonical form.

Solution: From Example 11 we know that (1) can be reduced to

$8y_1^2 + 2y_2^2$. ...... (2)

Now consider the coordinate transformation.

$z_1 = \sqrt{8}\, y_1$

$z_2 = \sqrt{2}\, y_2$

i.e., $Z = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix} Y$, where $Z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$.

This transformation, which is non-singular but not orthogonal, will convert (2) into

$z_1^2 + z_2^2$,

which is the required normal canonical form.

Example 14: Reduce the diagonal form

$2x_1^2 - 3x_2^2 - 7x_3^2$ into its normal canonical form.

Solution: Consider the transformation

$y_1 = \sqrt{2}x_1$,

$y_2 = \sqrt{3}x_2$

$y_3 = \sqrt{7}x_3$

i.e., $Y = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{7} \end{bmatrix} X$.

This will convert the given diagonal form into

$y_1^2 - y_2^2 - y_3^2$,

which is the required normal canonical form.

Try the following exercises now.

E E22) Reduce the following quadratic forms to their normal canonical forms.

a) $8x^2 - 4xy + 5y^2$

b) $2y^2 - 2yz + 2zx - 2xy$

E E23) Show that the rank of a normal canonical form is the number of non-zero terms in its expression.

E E24) Show that a quadratic form and its normal canonical reduction have the same rank.

In view of the above exercises a normal canonical reduction of a quadratic form of rank r has the form

$y_1^2 + \cdots\cdots + y_p^2 - y_{p+1}^2 - \cdots\cdots - y_r^2$,

where p is the number of positive terms in the reduced form.

But is a normal canonical reduction of a quadratic form unique? In other words, is the number of positive terms in a normal canonical reduction of a quadratic form uniquely determined? We answer this question in the following theorem, due to the English mathematician J.J. Sylvester (1814 - 1897).

Theorem 7 (Sylvester): The number of positive terms in a normal canonical reduction of a quadratic form is uniquely determined. Consequently, a quadratic form of rank r has a unique normal canonical reduction

$y_1^2 + \cdots\cdots + y_p^2 - y_{p+1}^2 - \cdots\cdots - y_r^2$

Proof: Let Q be a quadratic form of order n and rank r. Let $\{u_1, \ldots\ldots, u_n\}$ be a basis of R^n in which Q is represented by

$$Q(X) = x_1^2 + \ldots\ldots + x_p^2 - x_{p+1}^2 - \ldots\ldots - x_r^2, \qquad \ldots\ldots (1)$$

where $X = \sum_{i=1}^{n} x_i u_i$.

Let $\{v_1, \ldots\ldots\ldots, v_n\}$ be another basis of R^n in which Q is represented by

$$Q(Y) = y_1^2 + \ldots\ldots + y_{p'}^2 - y_{p'+1}^2 - \ldots\ldots - y_r^2, \qquad \ldots\ldots (2)$$

where $Y = \sum_{i=1}^{n} y_i v_i$.

Thus, (1) and (2) are both normal canonical reductions of Q, in which the number of positive terms are p and p′, respectively. To prove the theorem we have to prove that p = p′. Let U and V be the subspaces of R^n generated by $\{u_1, \ldots\ldots, u_p\}$ and $\{v_{p'+1}, \ldots\ldots, v_n\}$, respectively.

Thus, dim U = p and dim V = n − p′. We will show that $U \cap V = \{0\}$.

Suppose $U \cap V \neq \{0\}$. Let $0 \neq u \in U \cap V$.

Now, since $u \in U$ and $u \neq 0$, we have

$u = a_1 u_1 + \ldots\ldots + a_p u_p$, $a_i \in R\ \forall\ i$, where $a_i \neq 0$ for some i.

Therefore, from (1)

$$Q(u) = a_1^2 + \cdots + a_p^2 > 0 \qquad \ldots\ldots (3)$$

Also, since $u \in V$, we have

$u = b_{p'+1} v_{p'+1} + \ldots\ldots + b_n v_n$, $b_i \in R\ \forall\ i$, $b_i \neq 0$ for some i.

$\therefore$, from (2) we get $Q(u) = -b_{p'+1}^2 - \ldots\ldots - b_r^2 \leq 0 \qquad \ldots\ldots (4)$

(3) and (4) bring us to a contradiction. $\therefore$, our supposition must be wrong.

$\therefore U \cap V = \{0\}$.

At this stage, recall from Unit 3 that

dim U + dim V − dim $(U \cap V)$ = dim (U + V),

Therefore,

p + n − p′ = dim (U + V) ≤ dim (R^n) = n, as $U + V \subseteq R^n$.

$\Rightarrow p + n - p' \leq n$

$\Rightarrow p \le p'$

Interchanging the roles of p and p' in the above argument, we get

$p' \le p$ ...... (6)

(5) and (6) show that p = p', which proves the theorem.

By Theorem 1 and Sylvester's theorem the rank r and number p remain unchanged under a change of basis, i.e., under a non-singular transformation. Hence, the number 2p − r also remains unchanged.

Definition: The signature of a quadratic form is defined to be
(the number of positive terms) − (the number of negative terms) appearing in its normal canonical reduction. It is denoted by the letter s.

Thus, s = p − (r − p) = 2p − r.

For example, for the form in Example 13, we have p = 2, r = 2 and s = 2. For the form in Example 14, p = 1, r = 3, s = −1.

E E25) Find the rank and signature of the quadratic forms given in E 22.

The rank and the signature completely determine the normal canonical reduction. Also, any two quadratic forms having the same normal canonical reduction will be equivalent. We can, therefore, state the following result.

Theorem 8: Two quadratic forms are equivalent if and only if they have the same rank and signature.

In Section 14.3 we said that there is a one-to-one correspondence between the set of all symmetric matrices of order n and the set of quadratic forms of order n. So we can expect Sylvester's theorem to have a matrix interpretation. This is as follows:

A symmetric matrix of order n and rank r is equivalent to a unique diagonal matrix of the type

$$\begin{bmatrix} I_p & 0 & 0 \\ 0 & -I_{r-p} & 0 \\ 0 & 0 & 0_{n-r \times n-r} \end{bmatrix}.$$

And now we end the unit by briefly recalling what we have done in it.

14.8 SUMMARY

In this unit all the spaces considered are over the field **R**. In it we have covered the following points.

1) A homogeneous polynomial of degree two is called a quadratic form. Its order is the number of variables occurring in its expression.
2) Each quadratic form can be uniquely expressed as X'AX, where A is a unique symmetric matrix and is called the matrix of the quadratic form.
3) There is a one-to-one correspondence between the set of real symmetric n × n matrices and the set of real quadratic forms of order n.
4) Two quadratic forms are called equivalent (respectively, orthogonally equivalent) if their matrices are congruent (respectively, orthogonally similar). Two equivalent

(respectively, orthogonally equivalent) quadratic forms convert into each other by a suitable change of basis.

5) The rank of a quadratic form is defined to be the rank of its matrix.

6) A quadratic form X^tAX of rank r is orthogonally equivalent to a unique diagonal form

$$\lambda_1 y_1^2 + \ldots + \lambda_r y_r^2, \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r,$$

called its orthogonal canonical reduction, where $\lambda_1, \ldots, \lambda_r$ are the non-zero eigenvalues of A.

7) A quadratic form of rank r is equivalent to a unique diagonal form

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2,$$

called its normal canonical reduction. Here the number p is uniquely determined (Sylvester's theorem). The number 2p – r is called the signature of the quadratic form.

14.9 SOLUTIONS/ANSWERS

E1) There are plenty of possible answers. We give one each.

a) $x^2 + 1$, b) x^3.

E2) Only (a).

E3) a) $k = 0$, otherwise the polynomial is of degree 3.

b) $k = 2$

c) $k = 4$.

E4) The first three will be quadratic forms, if they are non-zero. Q_1Q_2 will be of degree 4. Q_1/Q_2 will also not be quadratic; in fact, it may not even be a polynomial.

E5) For example, the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}$ gives us the quadratic form

$$X^t \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} X = x^2 + 2xy + 3y^2.$$

E6) a) $\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}$, b) $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, c) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, d) $\begin{bmatrix} p & q/2 \\ q/2 & r \end{bmatrix}$

E7) a) $[x\ y\ z] \begin{bmatrix} 7 & -1 & -10 \\ -1 & 7 & 10 \\ -10 & 10 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

b) $[x_1\ x_2] \begin{bmatrix} 1 & \frac{-1}{2} \\ \frac{-1}{2} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

c) $[x_1\ x_2\ x_3] \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

d) $[x\ y\ z] \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

E8) a) $ax^2 + by^2 + cz^2 + 2hxy + 2gxz + 2fyz$

b) $x^2 - y^2$

c) $4x^2 - \sqrt{2}y^2 - z^2$

E9) $X^tAX = 4x_1^2 + x_2^2 + 4x_4^2 + 2x_1x_2 + 3x_1x_3 + 2x_1x_4 + 6x_2x_3 + x_3x_4$.

$$A' = \begin{bmatrix} 4 & 1 & 3/2 & 1 \\ 1 & 1 & 3 & 0 \\ 3/2 & 3 & 0 & 1/2 \\ 1 & 0 & 1/2 & 4 \end{bmatrix}$$

E10) Since its columns are not orthonormal, it is not orthogonal.

E11) Now X = PY, where $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. This is also not orthogonal, since its columns are not orthonormal.

Now $Q(Y) = Y^t (P^tAP)Y$, where

$$P^tAP = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

$\therefore Q(Y) = y_1^2 + 3y_2^2$.

E12) $A = \begin{bmatrix} 7 & 26 \\ 26 & -32 \end{bmatrix}$

The coordinate transformation corresponding to the change from B_1 to B_2 is given by the matrix $P = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix}$. $\therefore$, the matrix of the form will now be

$$P^tAP = \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 7 & 26 \\ 26 & -32 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -2 \end{bmatrix} = \begin{bmatrix} 100 & 0 \\ 0 & -225 \end{bmatrix}$$

$\therefore$, the quadratic form will now be expressed as $100x'^2 - 225y'^2$.

E13) Let the coordinates of a vector be $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$

with respect to the bases B and B', respectively. Then the coordinate transformation is given by

$$X = \begin{bmatrix} -2 & 3 & 6 \\ 6 & -2 & 3 \\ 3 & 6 & -2 \end{bmatrix} Y.$$

$$\Rightarrow x_1 = -2y_1 + 3y_2 + 6y_3$$
$$x_2 = 6y_1 - 2y_2 + 3y_3$$
$$x_3 = 3y_1 + 6y_2 - 2y_3$$

is the required coordinate transformation.

E14) The rank of the quadratic form $a_1x_1^2 + \ldots + a_nx_n^2$

= the rank of the matrix diag $(a_1, \ldots, a_n)$

= number of non-zero a_i's

= number of non-zero terms in the expression of the quadratic form.

E15) a) The rank of the form = rank of $\begin{bmatrix} 5 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 7 \end{bmatrix}$ = 3, since its determinant rank is 3.

b) rank (Q) = rank of $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ = 1, since its row-reduced echelon form is $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

c) rank (Q) = rank of $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$ = 3, since its determinant is non-zero.

d) rank (Q) = rank of $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ = 2, using the determinant rank method.

E16) Congruence is

i) reflexive : $A = I^tAI$

ii) symmetric: If $A = P^tBP$, then $B = (P^{-1})^t A (P^{-1})$.

iii) transitive: If $A = P^tBP$ and $B = R^tCR$ for some invertible matrices P and R, then $A = (RP)^t C(RP)$, and RP is an invertible matrix.

$\therefore$, congruence is an equivalence relation.

E16) Orthogonal similarity is

i) reflexive: $A = I^tAI$, and I is an orthogonal matrix.

ii) symmetric: If P is an orthogonal matrix such that $A = P^tBP$, then $B = (P^{-1})^t A(P^{-1})$, and P^{-1} is also an orthogonal matrix.

iii) transitive: $A = P^tBP$ and $B = R^tCR \Rightarrow A = (RP)^tC(RP)$.
Also P orthogonal and R orthogonal $\Rightarrow$ RP orthogonal.

$\therefore$ orthogonal similarity is an equivalence relation.

E17) The required transformation is $X = PY$, where $P = [-U_1 \ -U_2]$,

i.e., $x_1 = \frac{1}{\sqrt{2}}(y_1 - y_2)$

$x_2 = \frac{-1}{\sqrt{2}}(y_1 + y_2)$

E18) a) The matrix of the form is $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Its eigenvalues are 3 and -1. $\therefore$, the given form is equivalent to $3x_1^2 - y_1^2$. Normalised eigenvectors corresponding to 3 and -1 are $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}$, respectively. $\therefore$, they form a set of principal axes of the form. Remember that the principal axes are not unique.

b) Its orthogonal canonical form is $9x_1^2 + 4y_1^2$.

A set of principal axes is $\left\{\begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}\right\}$.

c) Its orthogonal canonical reduction is $4y_1^2 + 4y_2^2 - 2y_3^2$.

Eigenvectors corresponding to the eigenvalue 4 are given by

$$\begin{bmatrix} 0 & 2 & 2 \\ 2 & 3 & -1 \\ 2 & -1 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 4\begin{bmatrix} x \\ y \\ z \end{bmatrix} \Rightarrow 2x - y - z = 0.$$

$\therefore$, two linearly independent orthonormal eigenvectors corresponding to 4 can be obtained by putting $x = 0$ and $y = 0$ respectively, in this equation. So we get

$\begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{5} \\ 0 \\ 2/\sqrt{5} \end{bmatrix}$ as the required vectors.

Also, corresponding to the eigenvalue -2, we get a normalised eigenvector,

$\begin{bmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.

$\therefore$, a set of principal axes is

$$\left\{\begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ \frac{2}{\sqrt{5}} \end{bmatrix}, \begin{bmatrix} \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{bmatrix}\right\}.$$

E19) Any two forms are orthogonally equivalent iff they have the same orthogonal canonical forms as given in Theorem 5. $\therefore$, their matrices should have the same eigenvalues (including repetitions).

Now, the eigenvalues of the matrices in (a) and (c) are 12, 12 and -6. $\therefore$, the forms in (a) and (c) are orthogonally equivalent. The matrix of the form in (b) has eigenvalues 9, 9, -9. $\therefore$, it is not orthogonally equivalent to the others.

E20) Both the forms have the same diagonal form, as given in Theorem 5, namely $x'^2 + y'^2 - 2z'^2$. If

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = P \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \text{ and } \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = Q \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}, \text{ then}$$

PQ^{-1} will transform the first to the second, and

$PQ^{-1} = PQ^t$, since Q is orthogonal.

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

E21) The transformation (3) is given by Y = PZ, where

$$P = \text{diag}\left(\frac{1}{\sqrt{|\lambda_1|}}, \ldots\ldots, \frac{1}{\sqrt{|\lambda_r|}}, 1, \ldots\ldots, 1\right).$$

This matrix is orthogonal provided $PP^t = I$, i.e., $|\lambda_i| = 1 \ \forall \ i = 1, \ldots\ldots, r$, i.e. $\lambda_i = 1$ or $-1 \ \forall \ i = 1, \ldots\ldots, r$

E22) a) First obtain the orthogonal canonical form $9x_1^2 + 4y_1^2$. Then obtain its normal canonical form $x_2^2 + y_2^2$.

b) $x_1^2 - y_1^2$ is the normal canonical form.

E23) The rank of any diagonal form is the number of non-zero terms in its expression.

E24) Since the normal canonical reduction is obtained by non-singular transformations, the rank remains unchanged.

E25) a) rank = 2, signature = $2 \times 2 - 2 = 2$.

b) rank = 2, signature = $2 \times 1 - 2 = 0$.

UNIT 15 CONICS

Structure

15.1 INTRODUCTION

In Unit 14 you have studied about real quadratic forms of any order n. This unit is only a geometric extension of the previous one. In it we shall confine ourselves to the two-dimensional case.

Circles, parabolas, hyperbolas and ellipses are curves which we come across quite often. The ancient Greeks studied these curves and named them conic sections, since they could be obtained by taking a plane section of a right circular double cone (Fig. 1). However, from the analytic viewpoint, the Greek definition of conics, as sections of a cone, is not particularly useful. We shall consider a conic to be a curve which can be represented by an equation of second degree.

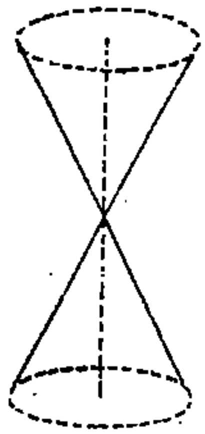


Fig. 1: Right circular double cone

After defining conics, we shall list the different types of standard conics. Then we shall study the ellipse, the hyperbola and the parabola in detail. In the last section we will look at one of the basic problems of plane analytic geometry that deals with conics—how to obtain a rectangular coordinate system in which the equation of a given conic takes the standard form.

Before going further, we suggest that you revise Unit 14

Objectives

After reading this unit, you should be able to

- recognise different types of conics and their standard equations;
- reduce a general equation of second degree to one of the standard forms of conics;
- trace a conic whose standard equation is given.

15.2 DEFINITIONS AND EQUATIONS

You have come across polynomials in several variables already. We will consider the curves that represent polynomials of degree two, in two variables.

15.2.1 What is a Conic?

Let us go back to Sec. 14.2, where we told you that the general equation of second degree in $\mathbb{R}^2$ is

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0, \qquad \ldots\ldots (1)$$

where a, h, b, g, f and c are real constants, of which at least one of a, h, b is non-zero. Note that if a, h, b are all zero, then (1) will become an equation of first degree, and hence, will represent a straight line.

Now, (1) represents a curve in R^2. We call this curve a conic. Let us make some formal definitions now.

Definitions: The set of points of R^2 whose coordinates satisfy an equation of second degree is called a **conic**.

It may happen that there is no point of R^2 that satisfies a given equation of second degree. (For example, no point of R^2 satisfies the equation $x^2 + y^2 = -1$.) In such a case we say that the conic represented by the equation is an **imaginary conic.**

Let us look at some examples.

Example 1: Investigate the nature of the conic given by

$x^2 + y^2 = a, a \in R$. ...... (2)

Solution: There are three cases to consider depending on the sign of a : $a < 0$, $a = 0$, $a > 0$.

Case 1: If $a < 0$, then no real values of x and y will satisfy (2), and therefore, the conic represented by (2) will be imaginary.

Case 2: If $a = 0$, then the only real solution of (2) is $x = 0$ and $y = 0$. Hence, the conic represented by (2) will consist of just one point, i.e., (0, 0).

A conic consisting of only one point is called a **point conic.**

Case 3: If $a > 0$, then $\sqrt{a} \in R$ and $a = (\sqrt{a})^2$. $\therefore$, a point (x, y) will satisfy (2) if and only if the distance of (x, y) from the origin is $\sqrt{a}$. Hence, the conic represented by (2) will be a circle of radius $\sqrt{a}$ and centre (0, 0).

Example 2: Find the nature of the conic represented by

$2x^2 - xy - 3x = 0$. ...... (3)

Solution: Equation (3) can be written as

$x(2x - y - 3) = 0$.

A first degree equation in R^2 represents a straight line.

This shows that a point (x, y) will satisfy (3) if it satisfies $x = 0$ or $2x - y - 3 = 0$. Therefore, we see that the points satisfying (3) are points of the lines $x = 0$ and $2x - y - 3 = 0$. $\therefore$, the conic consists of a pair of straight lines.

The examples above show that a circle, a point and a pair of straight lines are conics.

Try the following exercises now.

E E1) Find equations of second degree which will represent a pair of
(a) parallel lines, (b) coincident lines.

(**Hint**: Remember that parallel lines have the same slope.)

E E2) Find the nature of the conics represented by the following equations.

a) $x^2 - 2xy + y^2 = 0$

b) $4x^2 - 9x + 2 = 0$

c) $x^2 = 0$

d) $xy = 0$

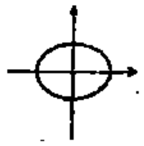
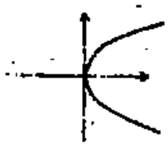
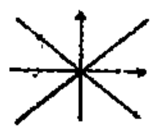
In the examples and exercises that you have done so far, you have dealt with simple second degree equations. These and other simple forms are what we will discuss now.

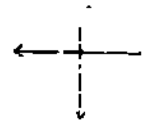
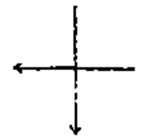
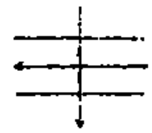
15.2.2 Standard Equations of Conics

Did you notice that we have not given any examples of conics like $x^2 + 5xy + y^2 + 2x - 6y + 10 = 0$ so far? We will do so in Sec. 15.6. And then you will see that we can always choose a coordinate system so that the equation of the conic in this system is in the "simplest" form, that is, it has as few terms as possible. Such a form is called the **standard equation of the conic**. In this sub-section we shall discuss this form.

There are several types of standard conics to which a general quadratic equation can be reduced. The classification is made on the basis of the coefficients of the various terms and the constant term appearing in the equation. In Table 1 we list different types of real conics along with their standard equations.

Table 1: Standard Forms of Conics

Conic	Standard Equation	Sketch
Ellipse	$x^2/a^2 + y^2/b^2 = 1$, a, b > 0	
Circle	$x^2 + y^2 = a^2$, $a \neq 0$	
Hyperbola	$x^2/a^2 - y^2/b^2 = 1$, a, b > 0	
Parabola	$y^2 = 4px$, p > 0	
Pair of intersecting lines	$x^2/a^2 - y^2/b^2 = 0$ a, $b \neq 0$	

Pair of parallel lines	$y^2 = a^2, a \neq 0$	
Pair of coincident lines	$y^2 = 0$	
Point conic	$x^2/a^2 + y^2/b^2 = 0, a, b \neq 0$	

From the standard equations of conics that we have listed in Table 1, we can obtain other equally simple equations by the following two methods.

i) **Interchanging the role of the axis:** We apply the orthogonal transformation

$$\left.\begin{array}{l} x = Y \\ y = X \end{array}\right\} \quad \ldots\ldots (1)$$

to the conic.

ii) **Reversing the direction of an axis:** For example, the direction of the x-axis can be reversed by applying the orthogonal transformation

$$\left.\begin{array}{l} x = -X \\ y = Y \end{array}\right\} \quad \ldots\ldots (2)$$

to the conic.

Similarly, we can reverse the direction of the y-axis by applying the orthogonal transformation $x = X, y = -Y$.

Let us illustrate the above discussion.

Example 3: Consider the standard equation $y^2 = 4px$ $(p > 0)$ of a parabola. What are the different forms of this equation that we can obtain under transformations (1) and (2)?

Solution: If we interchange the x and y axes, the given equation will transform to

$X^2 = 4pY, p > 0.$

To apply (2) we replace x by $-X$ and y by Y. Then the given equation will transform to

$Y^2 = -4pX, p > 0$

All three equations represent the same parabola with respect to different coordinate systems.

Try the following exercises now.

E E3) What are the different forms of the equation of the circle $x^2 + y^2 = a^2$ that we get on applying the transformations (1) and (2) given above?

Let us now study some of the conics in detail. In the following sections we will describe ellipses, hyperbolas, parabolas and other conics. As we go along we will also pictorially show you how conics occur as plane sections of a right circular double cone.

Before starting these sections you may like to recall what you studied about curve tracing in **Block 2 of the Calculus course.**

15.3 ELLIPSE

In the Foundation Course in Science and Technology, you have already studied that any planet orbits the sun in an elliptical path. The sun is at a focus of these ellipses. In this section, you will see what exactly an ellipse is and study some of its geometrical properties. In Fig. 2 you can see why an ellipse is called a conic.

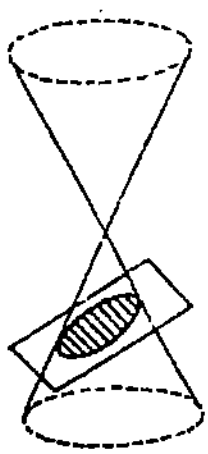


Fig. 2: Ellipse as a section of a double cone

15.3.1 Description

From Sec. 15.2 you know that the standard equation of an ellipse is

$$x^2/a^2 + y^2/b^2 = 1, \ a, b > 0 \qquad \ldots\ldots (1)$$

We may assume a > b. (If b > a, then we can interchange the x and y axes to arrive at the assumed case.) We want to trace the ellipse (1). For this purpose we start gathering information.

a) (1) is symmetric about the axes: If we replace x by (−x) or y by (−y) in (1), it remains unchanged. This shows that the ellipse is symmetric with respect to both the axes.

b) (1) is a central conic: If we replace both x and y by (−x) and (−y) in (1), it remains unchanged. Thus, the ellipse is symmetric with respect to the origin. Hence, (0, 0) is the centre of the ellipse.

(0, 0) is the centre of a conic $f(x, y) = 0$ if $f(-x, -y) = f(x, y)$. If a conic has a centre, it is called a central conic.

(a) and (b) tell us that it is enough to sketch the graph in the first quadrant only, i.e., for $x, y \geq 0$.

c) (1) is contained in the rectangle bounded by x = ±a and y = ±b: (1) can be written as

$x^2 = a^2 (1 - y^2/b^2)$.

This shows that there are no real values of x for $|y| > b$. Hence, the ellipse does not exist in the regions $y < -b$ and $y > b$. Similarly, writing the equation as

$y^2 = b^2 (1 - x^2/a^2)$,

we see that the ellipse does not exist in the regions given by $|x| > a$, i.e., for $x < -a$ and $x > a$.

d) (1) is bounded by the circle $x^2 + y^2 = a^2$.

If a point $P(x_1, y_1)$ lies on (1), then

$$\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2} = 1. \text{ Since } a \geq b, \text{ we get } \frac{y_1^2}{a^2} \leq \frac{y_1^2}{b^2}.$$

Therefore, $\dfrac{x_1^2 + y_1^2}{a^2} \leq \dfrac{x_1^2}{a^2} + \dfrac{y_1^2}{b^2} = 1$

i.e., $x_1^2 + y_1^2 \leq a^2$. This shows that P lies inside or on the circle $x^2 + y^2 = a^2$.

e) (1) intersects the coordinate axes in $(\pm a, 0)$ and $(0, \pm b)$.

f) The part of (1) in the first quadrant is given by

$$y = \frac{b}{a}\sqrt{a^2 - x^2}, \quad 0 \leq x \leq a \qquad \ldots\ldots (2)$$

$$\text{or} \quad x = \frac{a}{b}\sqrt{b^2 - y^2}, \quad 0 \leq y \leq b \qquad \ldots\ldots (3)$$

Here y is a continuous function of x, and it attains its maximum value b, at x = 0. As x increases continuously from 0 to a, y will continuously decrease from b to 0. From (2) above, y is a differentiable function of x over the interval [0, a]. The tangent at B(0, b) is y = b. From (3), x is a differentiable function of y over the interval [0, b], the tangent at A(a, 0) being x = a.

E4) Prove that the tangents at (a, 0) and (0, b) of the ellipse (1) are x = a and y = b, respectively.

From the above information the ellipse (1) will be represented by the curve in Fig. 3.

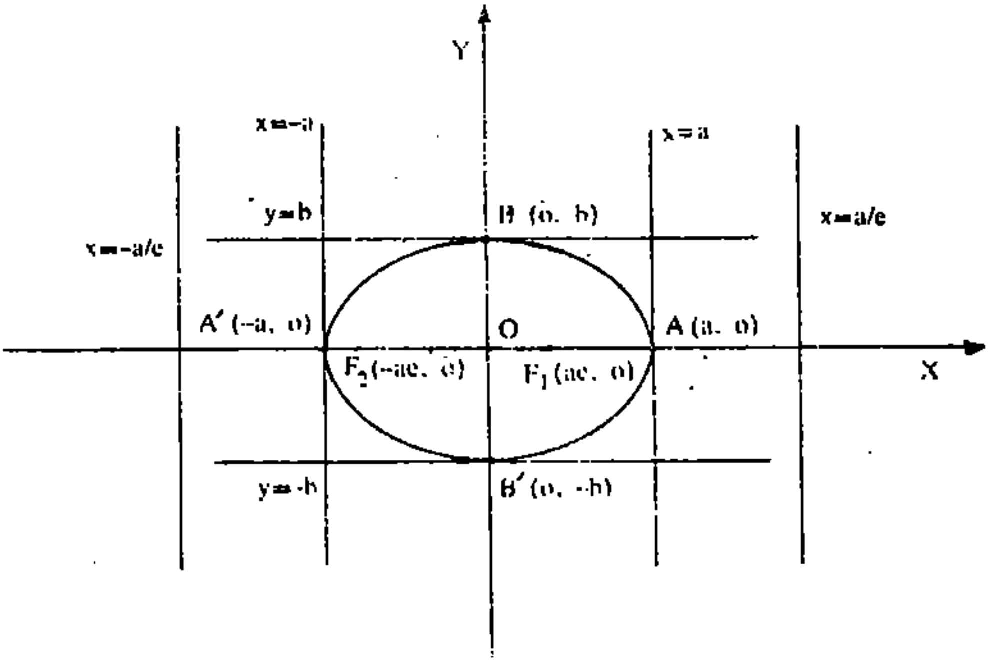


Fig. 3: The ellipse $x^2/a^2 + y^2/b^2 = 1$

The terms related to this ellipse are given below.

i) The points $(\pm a, 0)$ are called its **vertices**.

ii) A'A and B'B are called the **major and minor axes** of the ellipse, respectively. Their lengths are 2a and 2b, respectively.

These axes are the principal axes (ref. Sec. 14.6) of the ellipse. Can you see why? It is because they are given by the normalised eigenvectors $\begin{bmatrix}1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\end{bmatrix}$, of the form $x^2/a^2 + y^2/b^2$.

The major and minor axes of an ellipse are given by a set of normalised eigenvectors of its quadratic form.

iii) The positive real number e defined by
$a^2e^2 = a^2 - b^2$, is called the **eccentricity** of the ellipse. Note that $0 < e < 1$.

iv) The points (ae, 0) and (–ae, 0) are called the **foci** (plural of focus).

v) The line x = a/e is called the **directrix** (plural: directrices) corresponding to the focus (ae, 0). Similarly, x = –a/e is the directrix corresponding to (–ae, 0).

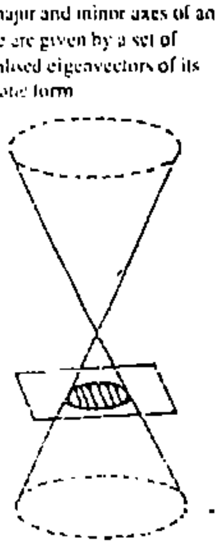


Fig. 4: Circle as a section of a cone

Note: If a = b the equation (1) reduces to $x^2 + y^2 = a^2$, which represents a circle of radius a (see Fig. 4). A circle is, thus, a special case of an ellipse.

We will study a circle in the following example.

Example 4: Find the eccentricity, foci and directrices of the circle $x^2 + y^2 = a^2$.

Solution: Since $x^2 + y^2 = a^2$ is a special case of (1) with b = a, we get e = 0. ∴ both the foci, $(\pm ae, 0)$ coincide at the origin, (0, 0). The two directrices $x = \pm a/e$ diverge to infinity as $e \to 0$, and do not exist in the real plane.

We have seen what happens if a = b in (1). But, what happens if b > a in (1)? The role of the major and minor axes will be interchanged and the terminology given for an ellipse will have to be suitably modified as follows:

i) the points (0, $\pm$ b) will be the vertices.

ii) B'B and A'A will be the major and minor axes, and their lengths will be 2b and 2a, respectively.

iii) the eccentricity e will be defined by

$b^2e^2 = b^2 - a^2$.

iv) the points (0, $\pm$ be) will be the foci. They will lie on the y-axis. Therefore, the major axis will lie along the y-axis.

v) The lines y = b/e and y = -b/e will be the directrices corresponding to the foci (0, be) and (0, -be) [illegible].

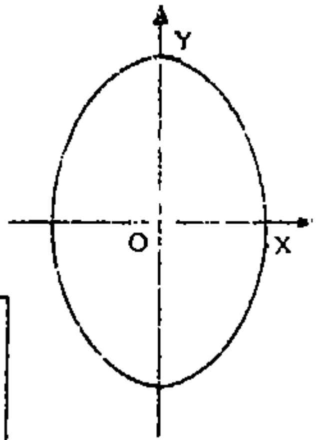


Fig. 5: The ellipse $x^2/4 + y^2/9 = 1$.

By [illegible] ellipse yourself. Try the following exercise.

E5) Find [illegible] directrices of the ellipse $9x^2 + 4y^2 = 36$ (see Fig. [illegible]

Now let us look closely at some properties of an ellipse.

15.3.2 Geometrical Properties

The ellipse has some very interesting geometrical properties. We shall study three important ones here.

Focus-directrix Property: The distance of any point of the ellipse from a focus is e times its distance from the corresponding directrix, where e is the eccentricity of the ellipse.

The eccentricity measures the ratio of the distance of a point on the curve from the focus and from the corresponding directrix.

Proof: Let $P(x_1, y_1)$ be a point on the ellipse $x^2/a^2 + y^2/b^2 = 1$, a > b (see Fig. 6). Let F_1 (ae, 0) be the focus under consideration. The directrix corresponding to F_1 is x = a/e. Let D be the foot of the perpendicular from P to the directrix x = a/e. Since P lies on the ellipse we have

$$x_1^2/a^2 + y_1^2/b^2 = 1 \Rightarrow b^2x_1^2 + a^2y_1^2 = a^2b^2$$
$$\Rightarrow (a^2 - a^2e^2)x_1^2 + a^2y_1^2 = a^2(a^2 - a^2e^2), \quad \text{since} \quad b^2 = a^2 - a^2e^2$$
$$\Rightarrow x_1^2 + y_1^2 + a^2e^2 = e^2x_1^2 + a^2.$$

Adding $-2aex_1$ on both sides, we get,

$$(x_1 - ae)^2 + y_1^2 = (ex_1 - a)^2$$
$$\Rightarrow (x_1 - ae)^2 + y_1^2 = e^2(x_1 - a/e)^2,$$

which is equivalent to

$PF_1^2 = e^2PD^2$, i.e., $PF_1 = e(PD)$,

which proves the statement for the focus F_1. For completing the proof, try E6.

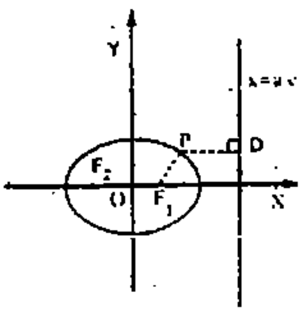


Fig. 6: The ellipse $x^2/a^2 + y^2/b^2 = 1$.

E6) Prove the focus–directrix property for the other focus F_2.

Another property that holds for ellipses is the

String Property: For each point P of the ellipse the sum of the distances of P from the two foci of the ellipse is the same, and is equal to the length of the major axis.

Proof: Let P be a point on the ellipse whose foci are F_1 and F_2 (see Fig. 6) Let D_1 and D_2 be the feet of the perpendiculars from P to the two directrices. Using the focus–directrix property, we get

$PF_1 + PF_2 = e(PD_1 + PD_2) = e(D_1D_2) = e(2a/e) = 2a,$

which proves the string property.

Fig. 7: Drawing an ellipse using string

You may wonder why this property is called the string property. It provides a mechanical method to construct an ellipse by using a string. Let us see what the method is.

A mechanical method for drawing an ellipse: Take a piece of string of length 2a and fix its end points at the points F_1 and F_2 ($F_1F_2 < 2a$) of a plane sheet of paper (see Fig. 7). Use the pencil point of a pencil to stretch the string into two segments. Now rotate the pencil point all around on the paper while sliding it along the string, making sure that the string is taut all the time. The curve traced will be an ellipse whose foci are F_1 and F_2, and the length of the major axis is 2a.

Fig. 8: Reflected wave property

E7) Use the method we have just given to draw an ellipse whose eccentricity is 0 and minor axis is 3 inches in length, on a piece of paper.

An ellipse has another important property which we shall state, but not prove in this course.

Reflected Wave Property: A ray of light (or sound, or any other type of wave) emitted from one focus of an ellipse is reflected back from its reflecting interior to the other focus (see Fig. 8).

A surface generated by revolving an ellipse about its major axis is called an ellipsoid.

An interesting consequence of this property is that rooms with an ellipsoidal ceiling have whispering galleries. A person standing at one focus of the ellipse can whisper so as to be heard by a person at the other focus, while the people in between cannot hear what is said.

Let us now study the hyperbola in detail.

15.4 HYPERBOLA

In this section we shall present the description and some geometrical properties of a hyperbola. See Fig. 9 for a representation of a hyperbola as a planar section of a double cone.

15.4.1 Description

From Table 1 you know that the standard equation of a hyperbola is

$$x^2/a^2 - y^2/b^2 = 1,\ a, b > 0. \qquad \ldots\ldots (1)$$

You can check that this is symmetric about both the axes, and hence about the origin. The origin is, therefore, the centre of the hyperbola. Thus, the hyperbola is a central conic.

The x-axis meets the hyperbola in $(\pm a, 0)$ while the y-axis does not meet it at all.

Due to symmetry about both the axes, it is enough to sketch the hyperbola in the first quadrant only, i.e., for $x, y \geq 0$. In this quadrant it is given by

$$y = b\sqrt{\frac{x^2}{a^2} - 1} \quad \left(\text{or } x = a\sqrt{\frac{y^2}{b^2} + 1}\right).$$

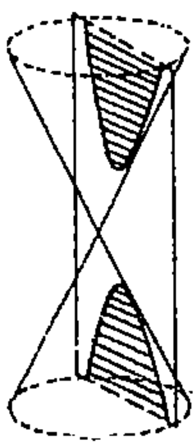


Fig. 9: Hyperbola as a section of a double cone

This provides the following information.

a) The hyperbola does not exist in the region $|x| < a$.

b) $y = 0$ for $x = a$.

c) y is a continuous function of x, which increases continuously from 0 to ∞ as x increases from 0 to ∞. The hyperbola, therefore, extends to infinity.

'∞' denotes infinity.

d) x is a differentiable function of y, and hence, a tangent can be drawn at each point of the hyperbola. The tangent at (a, 0) is parallel to the y-axis.

All this information allows us to sketch the hyperbola as in Fig. 10.

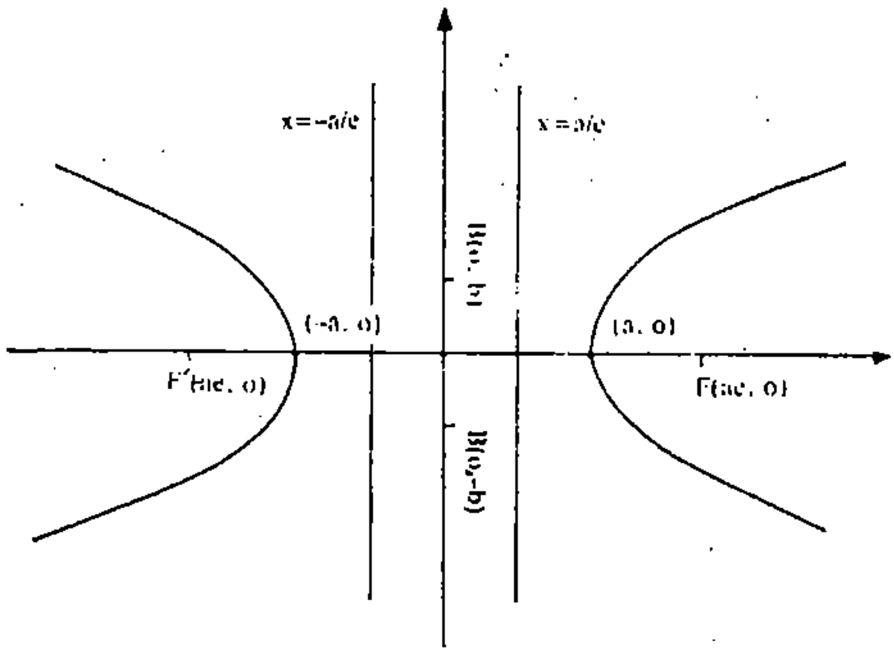


Fig. 10: The hyperbola $x^2/a^2 - y^2/b^2 = 1$

Can you see that the hyperbola consists of two branches? Of all the conics, this property is typical of hyperbolas only.

The terminology for the hyperbola is as follows:

i) The points $(\pm a, 0)$ are called its **vertices**.

ii) The line segment joining the vertices is called the **principal** (or **transversal**) **axis**, while the line segment joining B and B′ is called the **conjugate axis**. The length of the principal axis is 2a, while the length of the conjugate axis is 2b.

As in the case of an ellipse, these axes are in the direction of the normalised eigenvectors $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$, of the matrix of the form $x^2/a^2 - y^2/b^2$.

iii) The positive real number e, defined by

$a^2e^2 = a^2 + b^2$,

is called the **eccentricity** of the hyperbola.

Note that $e > 1$ in this case.

iv) The points $(\pm ae, 0)$ are the **foci** of the hyperbola.

v) The line $x = a/e$ (respectively, $x = -a/e$) is called the **directrix** corresponding to the focus (ae, 0) (respectively, (−ae, 0)).

Can you solve the following exercises now?

E E8) Find the vertices, eccentricity, foci and directrices of the hyperbola $9x^2 - 16y^2 = 144$.

Let us look at the geometry of a hyperbola now.

15.4.2 Geometrical Properties

A hyperbola has properties analogous to those of an ellipse. We discuss some important properties here.

Focus–directrix Property: The distance of any point of the hyperbola from either focus is e times its distance from the corresponding directrix.

Proof: We will start the proof and you can complete it! Let $P(x_1, y_1)$ be any point of the hyperbola $x^2/a^2 - y^2/b^2 = 1$, $a, b > 0$. Then $x_1^2 / a^2 - y_1^2 / b^2 = 1$. Consider the foci $F_1(ae, 0)$ and $F_2(-ae, 0)$. Now do E9.

E E9) Prove that $PF_1 = ePD$, where D = distance of P from the directrix $x = a/e$. Also show that $PF_2 = ePD'$, where D' = distance of P from the line $x = -a/e$.

So you have proved the focus–directrix property.

Corresponding to the string property of an ellipse we have the following property for a hyperbola.

String Property: For each point of a hyperbola the absolute value of the **difference** of its distances from the two foci is the same, and is equal to the length of the principal axis.

Proof: Let P be a point of the hyperbola whose foci are F_1 and F_2. Let D_1 and D_2 be the feet of the perpendiculars from P on the two directrices. Fig. 11 shows the two cases, when P is on one branch or the other.

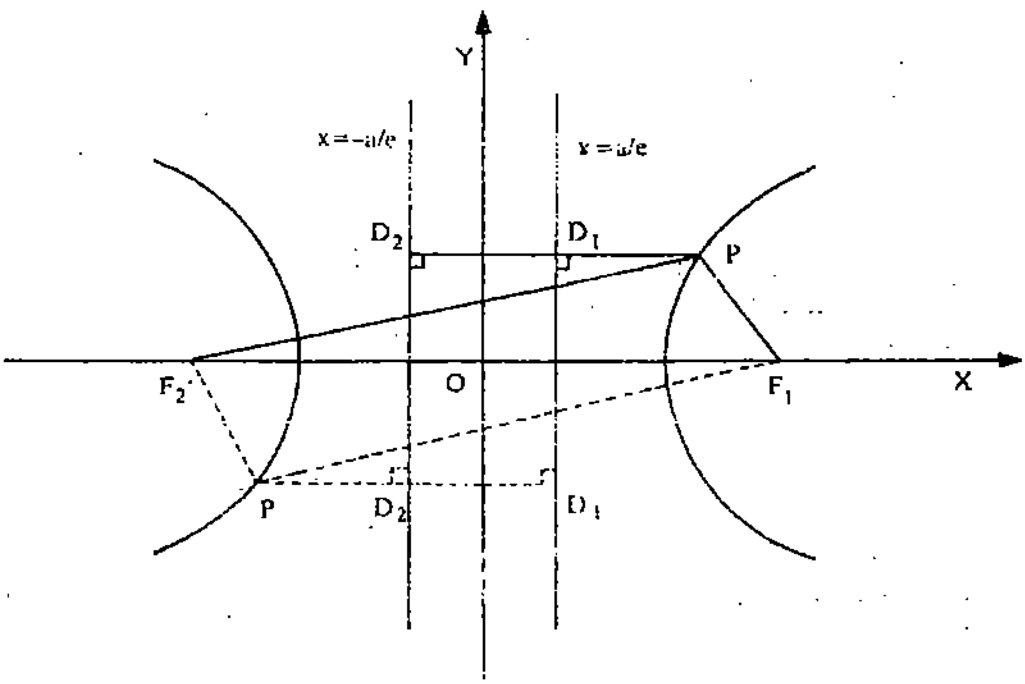


Fig. 11: String property for a hyperbola

From the focus–directrix property

$PF_1 = ePD_1$

$PF_2 = ePD_2$.

Hence,

$|PF_1 - PF_2| = e\,|PD_1 - PD_2| = e(D_1D_2) = e2a/e = 2a$, which proves the string property.

You must have noticed the similarity in the properties of an ellipse and a hyperbola. Sometimes an ellipse or a hyperbola is defined by the focus–directrix property, an ellipse being defined when $e < 1$, and a hyperbola when $e > 1$. What happens when $e = 1$? In other words, what is the locus of a point whose distance from a fixed point (a focus) is equal to its distance from a fixed line (a directrix)? We shall answer this question in the next section.

15.5 PARABOLA

Have you ever noticed the path of a projectile when it is acted upon by the force of gravity only? It is a parabola. In this section we will discuss parabolas in some detail. In Fig. 12 we show how it can be represented by a planar section of a cone.

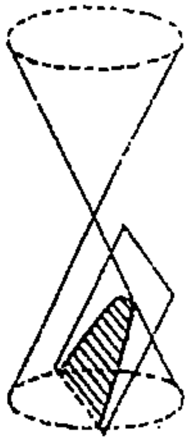


Fig. 12: Parabola as a section of a double cone

15.5.1 Description

Table 1 tells you that the standard equation of a parabola is $y^2 = 4px$, $p > 0$.

You can verify the following information about it, as you have done for an ellipse or a hyperbola.

a) It is symmetrical about the x-axis, and not about the y-axis.

   ∴, this is not a central conic.

b) For $x < 0$ there are no real values of y, and hence, this parabola does not exist in the second and third quadrants.

c) This parabola meets the axes only at the origin.

In view of (a) and (b), it is enough to sketch the parabola in the first quadrant only. The part of the parabola in the first quadrant is given by

$x = y^2/4p$ (or $y = 2\sqrt{px}$, $x \geq 0$).

x is a continuous and differentiable function of y, and hence, the tangent exists at each point.

The tangent at (0, 0) is the y-axis. As x increases continuously from 0 to ∞, y also increases from 0 to ∞. Hence the parabola is an infinite curve.

From the above information we draw the parabola in Fig. 13.

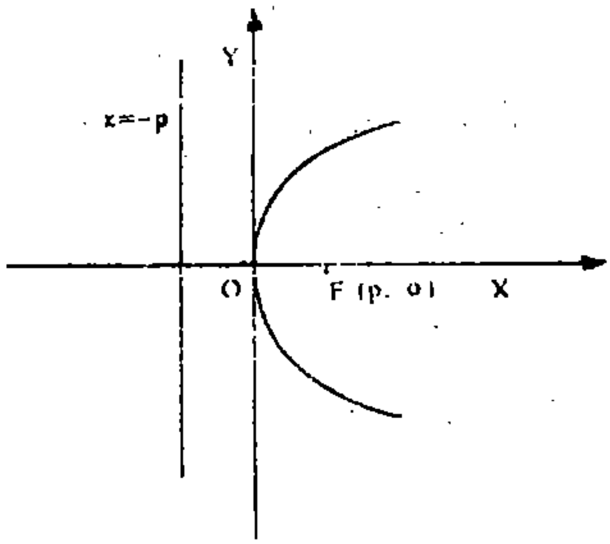


Fig. 13: The parabola $y^2 = 4px$

For the parabola given in Fig. 13,

i) the origin is called the vertex;

ii) the line of symmetry, i.e., the x-axis, is its axis.

iii) (p, 0) is the focus.

iv) the line $x = -p$ is the directrix.

You can use this knowledge to solve the following exercise.

E E10) Find the coordinates of the focus, and the equation of the directrix, of the parabola

a) $y^2 = 3x$, b) $x^2 = 4ay$, c) $y^2 = -4ax$.

Draw a rough sketch of these curves also.

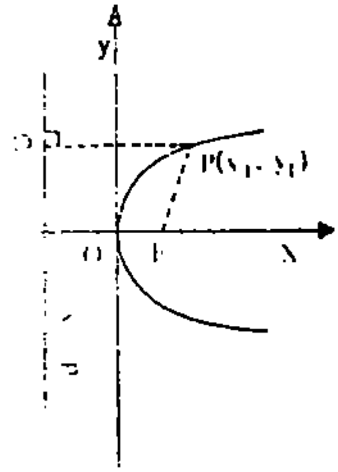


Fig. 14: PF = PD

We will now discuss the geometry of a parabola.

15.5.2 Geometrical Properties

We will talk about two geometrical properties of a parabola now.

Focus–directrix Property: Each point of a parabola is equidistant from the focus and the directrix of the parabola.

Proof: Let the parabola have standard equation $y^2 = 4px$. Then F(p, 0) is its focus. Let $P(x_1, y_1)$ be any point on the parabola (see Fig. 14). Then

$y_1^2 = 4px_1$.

Now

$$PF^2 = (x_1 - p)^2 + y_1^2 = (x_1 - p)^2 + 4px_1 = (x_1 + p)^2.$$

$$= PD^2$$

$$= (\text{distance of P from the directrix } x = -p)^2.$$

Hence, PF = PD, which proves the focus–directrix property.

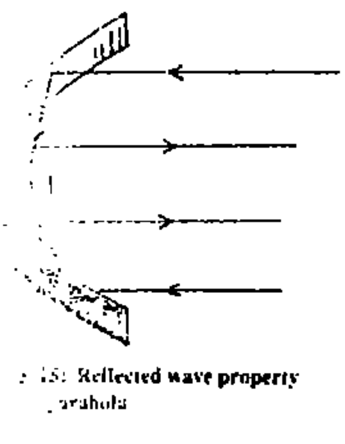


Fig. 15: Reflected wave property of parabola

Now we state (without proof) another important geometrical, as well as physical, property of a parabolic curve.

Reflected Wave Property: If a source of light (or sound, or any other type of wave) is placed at the focus of a parabola which has a reflecting surface (see Fig. 15), the rays that meet the reflecting surface of the parabola will be reflected parallel to the axis of the parabola. Conversely, the rays of light (or sound, or any other type of wave) entering parallel to the axis are reflected to converge at the focus.

As a consequence of this property a paraboloid surface is used in the headlight of cars, optical and radio telescopes, radars, etc.

A paraboloid is a surface generated by revolving a parabola about its axis.

The focus– directrix property is common to an ellipse, a hyperbola and a parabola. Each of them can be considered as a locus of a point whose distance from a fixed point (a focus) is a constant, e, times its distance from a fixed line (a directrix). The locus is an ellipse, parabola or hyperbola accordingly as $e < 1$, $e = 1$, $e > 1$. The focus– directrix property, therefore, unifies all these conics.

The ellipse, hyperbola and parabola are called non-degenerate conics.

What about the rest of the conics given in Table 1? They are all limiting cases of an ellipse, a hyperbola or a parabola

For example, the pair of intersecting lines $x^2 - k^2y^2 = 0$ is a limiting case of the hyperbola.

$x^2/a^2 - y^2/b^2 = 1$, a, $b > 0$ as $a \to 0$, $b \to 0$.

(Taking limits as $a \to 0$, $b \to 0$ such that $\lim\limits_{\substack{a\to 0\\ b\to 0}} a/b = k$ (finite), we get $x^2 - k^2y^2 = 0$.)

Similarly, the ellipse $x^2/a^2 + y^2/b^2 = 1$ degenerates into the pair of parallel lines given by $y^2 = b^2$, as $a \to \infty$.

So far you have studied quite a few conics. But you must be wondering about curves that are represented by the general equation of second degree.

We will now look at any conic and see how to reduce it to one of the standard forms given in Sec. 15.2.

15.6 THE GENERAL THEORY OF SECOND ORDER CURVES IN R^2

You know that the most general form of an equation of second degree is

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0, \qquad \ldots\ldots (1)$$

where a, h, b, g, f, $c \in$ **R** and a, h, b are not all zero.

We will see how to reduce this equation to standard form, that is, one of the forms listed in Table 1. You will see that the whole of this section will be devoted to using the following theorem.

Theorem 1: If the conic represented by (1) is not imaginary, then it is always possible to choose a rectangular coordinate system in which the equation (1) will reduce to one of the standard forms of conics.

We will give a rough outline of the proof of this theorem. The idea is to first reduce the quadratic form $ax^2 + 2hxy + by^2$ to the orthogonal canonical form $\lambda_1x_1^2 + \lambda_2y_1^2$, with $\lambda_1 \geq \lambda_2$ (cf. Sec. 14.5). Let this transformation be given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}.$$

On substituting these values of x and y in (1) we get a conic in x_1 and y_1. If this conic has any linear terms, we eliminate them by applying a translation of the form $x_1 = X + \alpha$, $y_1 = Y + \beta$, $\alpha, \beta \in$ **R**. We will choose α and β in such a manner that the linear terms are reduced to zero. Then our conic (1) will finally be transformed to one of the standard conics.

Our proof may seem vague to you. To understand the method of reduction consider the following examples.

Example 5: Reduce the conic $7x^2 - 8xy + y^2 = a$ to standard form. Hence, identify it.

Solution: The matrix of the quadratic form $7x^2 - 8xy + y^2$ is

$\begin{bmatrix} 7 & -4 \end{bmatrix}$

Its eigenvalues are 9 and −1. So, from Unit 14 (Theorem 5) you know that we can find an orthogonal transformation which will reduce $7x^2 - 8xy + y^2$ into $9X^2 - Y^2$. This transformation will reduce the given conic to

$9X^2 - Y^2 = a.$

The nature of this conic will depend on the value of a.

If $a = 0$, it will represent the pair of intersecting lines $3X - Y = 0$ and $3X + Y = 0$.

If $a \neq 0$, it will represent a hyperbola.

Example 6: Investigate the nature of the conic

$5x^2 - 6xy + 5y^2 + \sqrt{2}(x + y) = a.$

Solution: The second degree terms in the given equation are the same as in the quadratic form considered in Example 11 of Unit 14. The orthogonal coordinate transformation

$x = (1/\sqrt{2})(-y_1 + y_2)$

$y = (1/\sqrt{2})(y_1 + y_2)$

will convert $5x^2 - 6xy + 5y^2$ into $8y_1^2 + 2y_2^2$, and hence will transform the given equation into

$8y_1^2 + 2y_2^2 + (-y_1 + y_2 + y_1 + y_2) = a$.

i.e., $8y_1^2 + 2(y_2 + 1/2)^2 = a + 1/2$.

Now a translation of axes given by

$(y_1, y_2) \mapsto (X, Y - 1/2)$ will transform the above equation into $8X^2 + 2Y^2 = a + 1/2$, which is in standard form.

The nature of this conic will depend on the value of a. We have the following three cases:

Case 1: $a + 1/2 < 0$. In this case no real values of X and Y satisfy the conic, and hence the conic is imaginary.

Case 2: $a + 1/2 = 0$. In this case the conic is a point conic.

Case 3: $a + 1/2 > 0$. In this case the equation can be written as

$$\frac{X^2}{(2a+1)/16} + \frac{Y^2}{(2a+1)/4} = 1,$$

which represents an ellipse.

Note that we have used two successive transformations in Example 6 to convert the given equation into standard form. The first one was an orthogonal transformation. The second one was a translation. Both these transformations preserve the geometric nature of the curve. Thus, the given equation and its reduced form represent the same conic in the coordinate systems (x, y) and (X, Y), respectively.

Over here we would like to make the following remark.

Remark: When we apply an orthogonal transformation, what are we doing geometrically? We are simply rotating the axes. In fact, **orthogonal matrices correspond to rotations and reflections.**

In the following example you can see what a conic looks like before and after reduction to standard form.

Example 7: Let $a = 4$ in the equation considered in Example 6. Find the coordinate transformation that will convert it into standard form.

Solution: The composite of the two transformations in Example 6 is

$x = \frac{1}{\sqrt{2}}(-X + Y - 1/2)$

$y = \frac{1}{\sqrt{2}}(X + Y - 1/2)$,

which is the required coordinate transformation. Solving for X and Y we get

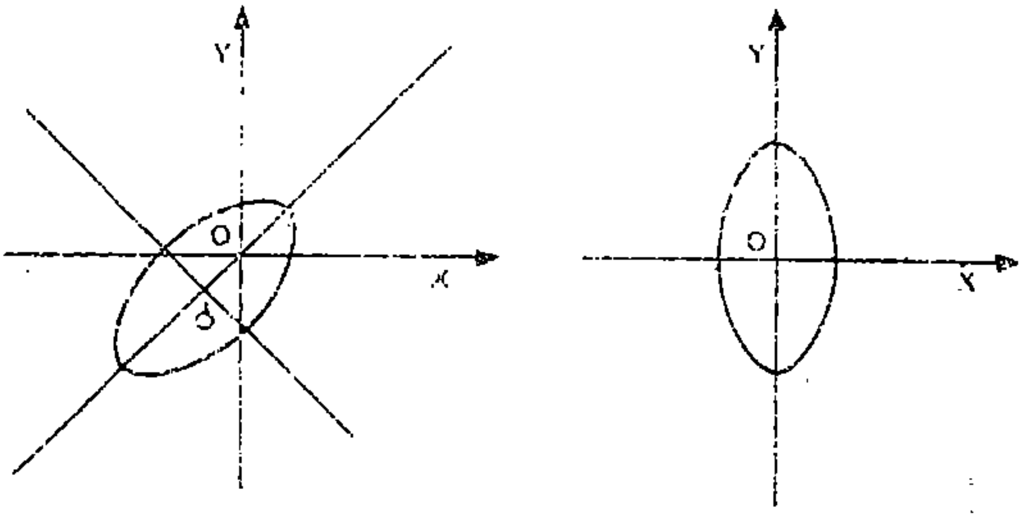
$X = \frac{y - x}{\sqrt{2}}$

$Y = \frac{y + x}{\sqrt{2}} + \frac{1}{2}$

For a = 4 the reduced equation becomes

$$\frac{X^2}{9/16} + \frac{Y^2}{9/4} = 1.$$

We give the sketch of the original equation in Fig. 16(a), and the sketch of the reduced equation in Fig. 16(b).



(a) (b)

Fig. 16: The ellipse $5x^2 - 6xy + 5y^2 + \sqrt{2}(x + y) = 4$
(a) before reduction, (b) after reduction.

So, you see, the shape and size of the conic remains unchanged under the transformations that we apply to reduce it to standard form.

Let us look at another example in which we identify a conic by reducing it to standard form.

Example 8: Find the nature of the conic

$x^2 + 2xy + y^2 - 6x - 2y + 4 = 0$

Solution: The matrix of the quadratic form $x^2 + 2xy + y^2$ is $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, whose eigenvalues are 2, 0. Normalised eigenvectors corresponding to the eigenvalues 2 and 0 are $(1/\sqrt{2},\ 1/\sqrt{2})$ and $(-1/\sqrt{2},\ 1/\sqrt{2})$, respectively. Hence, the coordinate transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

i.e., $x = (y_1 - y_2)/\sqrt{2}$, $y = (y_1 + y_2)/\sqrt{2}$,

will convert $x^2 + 2xy + y^2$ into $2y_1^2$, and the given equation into

$2y_1^2 - 3\sqrt{2}(y_1 - y_2) - \sqrt{2}(y_1 + y_2) + 4 = 0$

i.e., $(y_1 - \sqrt{2})^2 = -\sqrt{2}\, y_2$.

Now, we want to get rid of the linear terms. If we apply the translation

$y_1 - \sqrt{2} = X,\quad y_2 = Y,$

we can reduce the conic further into $X^2 = -\sqrt{2}Y$.

This represents a parabola. Hence, the given equation represents a parabola.

Let us formally write down what we have done in the various examples.

Step by step procedure for reducing a second degree equation in $\mathbf{R}^2$: Consider the second degree equation

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0 \quad \ldots\ldots (1)$$

Step 1: Use the method of Section 14.6 to reduce $ax^2 + 2hxy + by^2$ to $\lambda_1 y_1^2 + \lambda_2 y_2^2$ using an orthogonal transformation. This transformation will reduce (1) to

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + Ay_1 + By_2 + C = 0 \quad \ldots\ldots (2)$$

Step 2: Now use a suitable translation of axes $(y_1, y_2) \mapsto (X, Y)$ to eliminate the linear terms and reduce (2) into one of the standard forms. This will give the reduction of (1).

By now you must be wanting to try and reduce equations on your own. Try this exercise.

E E11) Reduce the following second degree equations to standard form. (Here $a \in \mathbf{R}$.) What is the type of conic they represent?

a) $x^2 + 4xy + y^2 = a$

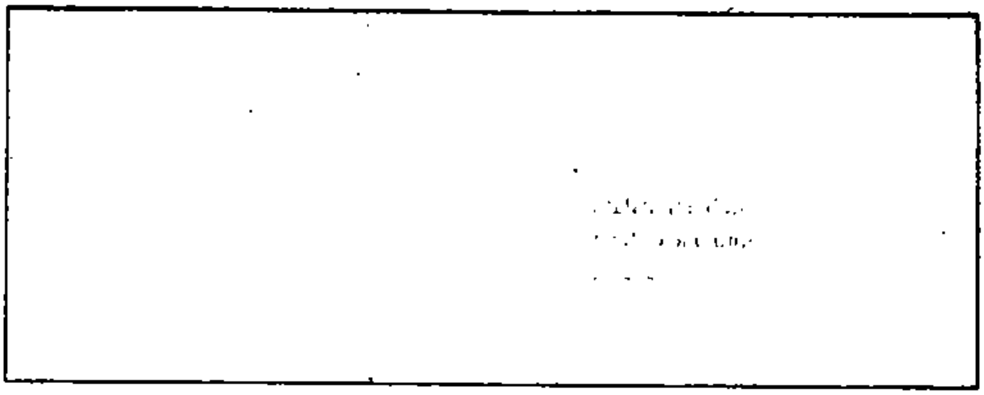
b) $8x^2 - 4xy + 5y^2 = a$

c) $3x^2 - 4xy = a$

d) $4x^2 - 4xy + y^2 = 1$

e) $16x^2 - 24xy + 9y^2 - 104x - 172y + 44 = 0$

f) $4x^2 - 4xy + y^2 - 12x + 6y + 9 = 0$



We end this unit with briefly mentioning what has been done in it.

15.7 SUMMARY

In this unit we have covered the following points.

1. A conic is defined to be the set of points in R^2 that satisfy an equation of second degree. Conics can be real or imaginary.
2. Real conics can be one of the following types:

   ellipse, circle, hyperbola, parabola, pair of straight lines, pair of parallel lines, pair of coincident lines, or a point. Their standard equations are listed in Table 1.
3. All these conics, except for a pair of parallel lines, can be obtained by taking a plane section of a right circular double cone.
4. An ellipse, a parabola and a hyperbola satisfy the focus-directrix property, i.e., the distance of any point P on them from a fixed point (a focus) is e (the eccentricity) times the distance of P from a fixed line (a directrix).
5. The ellipse and hyperbola have two foci and two corresponding directrices, while the parabola has one focus and one directrix.
6. $e = 1$, $e > 1$ or $e < 1$ accordingly as the conic is a parabola, a hyperbola or an ellipse.
7. An ellipse (a hyperbola) satisfies the string property, i.e., for each point P on the ellipse (hyperbola), the sum (absolute value of the difference) of the distances of P from the two foci is constant, and is equal to the length of the major (principal) axis.
8. The ellipse and parabola satisfy the reflected wave properties.
9. The ellipse, hyperbola and parabola are called non-degenerate conics. The rest of the conics can be obtained as limiting cases of the non-degenerate conics. The ellipse and hyperbola are non-degenerate conics with a unique centre, and hence, are called central conics.
10. Any second degree equation can be reduced to standard form by orthogonal transformations and translations.

15.8 SOLUTIONS/ANSWERS

E1) There can be many answers. We give the following:

a) $y = x + 1$ and $y = x - 1$ are a pair of parallel lines.

$\therefore \{y-(x+1)\}\{y-(x-1)\} = 0$ represents a pair of parallel lines.

b) $\{y-(x+1)\}^2 = 0$ represents a pair of lines, both of which are $y = x + 1$.

E2) a) $x^2 - 2xy + y^2 = 0 \Leftrightarrow (x-y)^2 = 0$. This represents the pair of coincident lines $x - y = 0$, i.e., $y = x$.

b) The equation represents the pair of parallel lines

$(x-2)\left(x-\frac{1}{4}\right) = 0$, i.e., $(x-2)(4x-1) = 0$.

c) The coincident lines $x = 0$, i.e., the y-axis.

d) The pair of lines $x = 0$ and $y = 0$, i.e., the y-axis and the x-axis.

NOTES